THE ANNALS OF MATHEMATICAL STATISTICS

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland.

Subscription Rates. Current issues are $12 per volume (four issues of one calendar year) in the U. S. and Canada, $10 per volume elsewhere. Single issues are $3.50. Back numbers are $12 per volume, $3.50 per issue. $200 for the first 25 volumes, $10 per additional volume purchased at the same time as Volumes I through XXV.

Communications concerning subscriptions, back numbers, payments of dues, etc. should be addressed to A. H. [illegible], Treasurer, Department of Statistics, Stanford University, Stanford, Calif.

Communications concerning membership, changes of address, etc. should be addressed to George E. Nicholson, Jr., Secretary, Department of Statistics, University of North Carolina, Chapel Hill, N. C. Changes of address which are to become effective for a given issue of the *Annals* should be reported to the Secretary on or before the 10th of the month preceding the month of issue.

Editorial Office. The RAND Corporation, 1700 Main St., Santa Monica, Calif., T. E. Harris, Editor.

Preparation of manuscripts. Manuscripts should be submitted to the editorial office. Each manuscript should be typewritten, double spaced, with wide margins at sides, top and bottom, and the original should be submitted with one additional copy, on paper that will take corrections. Dittoed or mimeographed papers are acceptable only if completely legible. Footnotes should be reduced to a minimum, and where possible replaced by remarks in the text, or a bibliography at the end of the paper; formulae in footnotes should be avoided. References should follow current *Annals* style.

Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed.

Authors are asked to keep in mind the typographical [illegible] of complicated [illegible]

(Continued on inside back cover)

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC., BALTIMORE, MD., U. S. A.

PÓLYA TYPE DISTRIBUTIONS, II[1]

By Samuel Karlin

Stanford University

In a previous publication a specific smoothing property characterizing a class of distributions which we called Pólya Type (P.T.) distributions was introduced [1]. Most of the standard distributions occurring in statistical practice are of Pólya Type. For this class of distributions many of the usual decision theoretic questions were analyzed. Explicitly in the case of the two action problem, complete classes of statistical procedures were characterized and Bayes and admissible procedures were also determined. This paper continues the further development of statistical applications for Pólya type distributions. We are still principally concerned with the two action problem. In a subsequent publication the n-action and estimation problem for P.T. distributions will be presented.

Our investigation is divided into three main parts. Part I describes some new characterizations of P.T. distribution. Attention is called to Lemma 3 which is very useful in establishing the fundamental variation diminishing properties of P.T. distributions as described in Theorem 3. Finally, Part I closes with two further results about the sums of two random variables one of which has a P.T. distribution.

In Part II we examine in detail many of the standard Neyman-Pearson concepts for the case when the underlying distributions are known to be Pólya Type. Representative topics treated include the principle of unbiasedness, envelope power functions, likelihood ratio tests, etc. Specifically, it is shown that in any testing problem uniformly most powerful unbiased tests always exist and in fact can easily be explicitly constructed. Although we deal here with the case where there is only a single free parameter, for many examples a problem involving several parameters can be reduced to that of one parameter by using the principle of similarity or the principle of invariance. At this point our theory can be directly applied. Another interesting consequence of the theory is the result that the likelihood ratio test for a composite hypothesis versus a composite alternative when the underlying family of distributions are of P.T. is an admissible test.

A general minimax theorem for the two action decision problem is developed in Part III. Explicitly the game defined by the usual risk function is shown to have a value (under very mild conditions imposed on the loss functions). Furthermore, the optimal strategies for both the statistician and nature are characterized. Specific attention is directed to the one and two-sided testing problems. A discussion of the computational job for obtaining the minimax strategies is also given.

Received November 18, 1955.

[1] Technical Report 1, prepared under contract Nonr-220(16) for the Office of Naval Research, Reference No. NR 042-995.

Although the three chapters are basically related, they may be read separately with only few references to other parts.

Finally I wish to express my gratitude to Rupert Miller for his help in the writing of the manuscript.

Part I. Definition and Properties of Pólya Type Distributions.

Sec. 1. Definitions and preliminaries.

Def. 1. A family of distributions $P(x, \omega)$

$$P(x, \omega) = \beta(\omega) \int_{-\infty}^{x} p(x, \omega)\, d\mu(x)$$

of a real random variable X depending on a real parameter ω is said to belong to the class $\mathcal{P}_n$ (Pólya Type n) if

$$\begin{vmatrix} p(x_1, \omega_1) & \cdots & p(x_1, \omega_m) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ p(x_m, \omega_1) & \cdots & p(x_m, \omega_m) \end{vmatrix} \geqq 0 \tag{1.1}$$

for every $1 \leqq m \leqq n$ and all $x_1 < x_2 < \cdots < x_m$, and $\omega_1 < \omega_2 < \cdots < \omega_m$. The family belongs strictly to $\mathcal{P}_n$ if strict inequality holds in (1.1). μ is a σ-finite measure on the real line and $p(x, \omega)$ is taken to be continuous in each variable. Most of the results can be easily extended to the case where we allow $p(x, \omega)$ to have a finite number of discontinuities of the first kind, in each variable separately.

If the family of distributions $P(x, \omega)$ belongs to $\mathcal{P}_n$ for every n, then we say that the family belongs to $\mathcal{P}_\infty$. If it belongs strictly to $\mathcal{P}_n$ for every n, then it belongs strictly to $\mathcal{P}_\infty$. We shall sometimes say that $p(x, \omega)$ is Pólya Type $n(\infty)$ if $P(x, \omega)$ belongs to $\mathcal{P}_n(\mathcal{P}_\infty)$.

For $n = 1, 2$ the conditions of being Pólya Type n reduce to familiar ones. p is Pólya Type 1 (strictly Pólya Type 1) if and only if $p(x, \omega) \geqq 0 (> 0)$ for all x and ω. p is Pólya Type 2 if and only if it has a monotone likelihood ratio, i.e., for every $x_1 < x_2$, $[p(x_1, \omega)]/[p(x_2, \omega)]$ is nonincreasing in ω. It is strictly Pólya Type 2 if and only if it has a strict monotone likelihood ratio, i.e., $[p(x_1, \omega]/[p(x_2, \omega)]$ is decreasing in ω for $x_1 < x_2$.

The distributions that can be classified as Pólya Type include almost all of the principal distributions occurring in statistical practice. The exponential family, the noncentral t, the noncentral F, and the noncentral chi-square distributions all belong strictly to $\mathcal{P}_\infty$. For a proof of this the reader is referred to [1]. Other examples are given in [2]. The most notable example of a density which is not Pólya Type is the Cauchy, i.e.,

$$p(x, \omega) = \frac{1}{\pi} \frac{1}{1 + (x - \omega)^2}.$$

Sec. 2. Some characterizations of Pólya Type distributions. This section will be devoted to presenting some alternative characterizations and some analytic properties of Pólya Type distributions. Theorem 3 and its corollaries should be carefully noted because the decision theory for Pólya Type distributions developed alternately in [1] is based almost entirely on this theorem.

THEOREM 1. *If p is Pólya type 2 and the derivatives involved exist everywhere, then*

$$\begin{vmatrix} p(x_1, \omega) & \frac{\partial}{\partial \omega} p(x_1, \omega) \\ p(x_2, \omega) & \frac{\partial}{\partial \omega} p(x_2, \omega) \end{vmatrix} \geqq 0 \tag{2.2}$$

for all ω and all $x_1 < x_2$, and

$$\begin{vmatrix} p(x, \omega) & \frac{\partial}{\partial \omega} p(x, \omega) \\ \frac{\partial}{\partial x} p(x, \omega) & \frac{\partial^2}{\partial x \partial \omega} p(x, \omega) \end{vmatrix} \geqq 0 \tag{2.3}$$

for all ω and all x. Conversely, if $p(x, \omega) > 0$ for all x and ω, (2.3) implies (2.2), which in turn implies that p is Pólya type 2. Strict inequality in (2.3) implies strict inequality in (2.2) and this implies that p is strictly Pólya Type 2.

REMARK. The requirement that $p(x, \omega) > 0$ in the converse theorem can be greatly relaxed by use of a device which will be fully explained in connection with Theorem 2 below.

PROOF. $p \in \mathcal{P}_2$ implies that for all $x_1 < x_2$ and $\omega_1 < \omega_2$

$$\frac{1}{\omega_2 - \omega_1} \begin{vmatrix} p(x_1, \omega_1) & p(x_1, \omega_2) \\ p(x_2, \omega_1) & p(x_2, \omega_2) \end{vmatrix} = \begin{vmatrix} p(x_1 \omega_1) & \frac{p(x_1, \omega_2) - p(x_1, \omega_1)}{\omega_2 - \omega_1} \\ p(x_2, \omega_1) & \frac{p(x_2, \omega_2) - p(x_2, \omega_1)}{\omega_2 - \omega_1} \end{vmatrix} \geqq 0.$$

The limit as ω_2 approaches ω_1 gives (2.2). Also, (2.3) is obtained from (2.2) analogous to the preceding by operating on columns.

The converse is established by showing that p has a monotone likelihood ratio. Indeed, (2.3) can be written as

$$[p(x, \omega)]^2 \frac{\partial}{\partial x} \left\{ \frac{\frac{\partial}{\partial \omega} p(x, \omega)}{p(x, \omega)} \right\} \geqq 0 \tag{2.4}$$

for all x and ω. This implies that $(\partial/\partial\omega)p(x, \omega)/p(x, \omega)$ is nondecreasing in x for all ω. This yields (2.2) which in turn implies

$$[p(x_1, \omega)^2] \frac{\partial}{\partial \omega} \left\{ \frac{p(x_2, \omega)}{p(x_1, \omega)} \right\} \geqq 0 \tag{2.5}$$

for all $x_1 < x_2$. (2.5) implies $[p(x_2, \omega)]/[p(x_1, \omega)]$ is nondecreasing in ω for all $x_1 < x_2$; i.e., p has a monotone likelihood ratio.

The strict converse is obtained by replacing $\geqq$ by $>$, nondecreasing by increasing, and monotone by strictly monotone in the preceding paragraph.

COROLLARY 1. *Suppose* $(\partial^2/\partial x \partial \omega)$ *log* $p(x, \omega)$ *exists and* p *belongs strictly to* $\mathcal{P}_1$. *Then* p *belongs to* $\mathcal{P}_2$ *if and only if*

$$\frac{\partial^2}{\partial x \partial \omega} \log p(x, \omega) \geqq 0$$

for all x *and* ω.

Our attention is now directed to consider a generalization of Theorem 1 for density functions which are Pólya Type of arbitrary degree.

THEOREM 2. *If* p *is Pólya Type* m *and all the derivatives involved exist everywhere, then*

$$\begin{vmatrix} p(x_1, \omega) & \frac{\partial}{\partial \omega} p(x_1, \omega) & \cdots & \frac{\partial^{n-1}}{\partial \omega^{n-1}} p(x_1, \omega) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ p(x_n, \omega) & \frac{\partial}{\partial \omega} p(x_n, \omega) & \cdots & \frac{\partial^{n-1}}{\partial \omega^{n-1}} p(x_n, \omega) \end{vmatrix} \geqq 0 \tag{2.6}$$

for all $n \leqq m$, ω, *and* $x_1 < x_2 < \cdots < x_n$, *and*

$$\begin{vmatrix} p(x, \omega) & \frac{\partial}{\partial \omega} p(x, \omega) & \cdots & \frac{\partial^{n-1}}{\partial \omega^{n-1}} p(x, \omega) \\ \frac{\partial}{\partial x} p(x, \omega) & \frac{\partial^2}{\partial x \partial \omega} p(x, \omega) & \cdots & \frac{\partial^n}{\partial x \partial \omega^{n-1}} p(x, \omega) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \frac{\partial^{n-1}}{\partial x^{n-1}} p(x, \omega) & \frac{\partial^n}{\partial x^{n-1} \partial \omega} p(x, \omega) & \cdots & \frac{\partial^{2n-2}}{\partial x^{n-1} \partial \omega^{n-1}} p(x, \omega) \end{vmatrix} \geqq 0 \tag{2.7}$$

for all $n \leqq m$, ω, *and* x. *Conversely, strict inequality in* (2.6) *for every* $1 \leqq n \leqq m$ *implies that* p *is strictly Pólya Type* m *and strict inequality in* (2.7) *for all* $1 \leqq n \leqq m$ *implies the same in* (2.6).

PROOF. We need the following lemma.

LEMMA 1. *If $f_1, f_2, \cdots, f_n$ are differentiable real-valued functions on the real line and $\xi_1 < \xi_2$, then there exists a $\bar{\xi}$, $\xi_1 < \bar{\xi} < \xi_2$, such that*

$$\begin{vmatrix} a_{11} & \cdots & a_{1j} & f_1(\xi_1) & f_1(\xi_2) & a_{1,j+3} & \cdots & a_{1n} \\ \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ a_{n1} & \cdots & a_{nj} & f_n(\xi_1) & f_n(\xi_2) & a_{n,j+3} & \cdots & a_{nn} \end{vmatrix}$$

$$= (\xi_2 - \xi_1) \begin{vmatrix} a_{11} & \cdots & a_{1j} & f_1(\xi_1) & f_1'(\bar{\xi}) & a_{1,j+3} & \cdots & a_{1n} \\ \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ a_{n1} & \cdots & a_{nj} & f_n(\xi_1) & f_n'(\bar{\xi}) & a_{1,j+3} & \cdots & a_{nn} \end{vmatrix},$$

where the a_{ij}'s are any real numbers.

The proof of this lemma is an easy application of the mean value theorem and will be omitted.

The proof of the first part of the theorem proceeds as follows. Let $p_i(\omega) = p(x_i, \omega)$ and $p_i^k(\omega) = (\partial^k/\partial\omega^k)p_i(\omega)$. Suppose n and $x_1 < x_2 < \cdots < x_n$ are given. For $\omega_1 < \omega_2 < \cdots < \omega_n$

$$0 \leqq \operatorname{sgn} \begin{vmatrix} p_1(\omega_1) & \cdots & p_1(\omega_n) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ p_n(\omega_1) & \cdots & p_n(\omega_n) \end{vmatrix} \tag{2.8}$$

$$= \operatorname{sgn} \begin{vmatrix} p_1(\omega_1) & p_1^1(\omega_2^1) & p_1^2(\omega_3^2) & p_1^3(\omega_4^3) & \cdots & p_1^{n-1}(\omega_n^{n-1}) \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ p_n(\omega_1) & p_n^1(\omega_2^1) & p_n^2(\omega_3^2) & p_n^3(\omega_4^3) & \cdots & p_n^{n-1}(\omega_n^{n-1}) \end{vmatrix},$$

where $\omega_{i-1}^{j-1} \leqq \omega_i^j \leqq \omega_i^{j-1}$. This equality is obtained by repeated application of Lemma 1. Sgn is the function which equals $+1$ if its argument is positive, -1 if its argument is negative, and 0 if its argument is zero. Letting $\omega_2 \to \omega_1, \omega_3 \to \omega_1, \cdots, \omega_n \to \omega_1$, the last determinant approaches the determinant in (2.6) and therefore (2.6) must hold. (2.7) is derived from (2.6) by applying the same

operations on the rows of the determinants in (2.6) as were applied to the columns of the first determinant in (2.8).

The proof of the converse of this theorem depends on Lemma 2 below. Lemmas 3 and 4 which will be needed subsequently in proving Theorem 3 are also included at this point because of the similarity of their proofs with the proof of Lemma 2.

LEMMA 2. *If all the derivatives involved exist and are continuous and*

$$\begin{vmatrix} p_1(w) & p_1^1(w) & \cdots & p_n^{n-1}(w) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ p_n(w) & p_n^1(w) & \cdots & p_n^{n-1}(w) \end{vmatrix} > 0 \tag{2.9}$$

for all w and $n \leqq m$ where p_i^j is the jth derivative of p_i, then

$$\begin{vmatrix} p_1(w_1) & \cdots & p_1(w_n) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ p_n(w_1) & \cdots & p_n(w_n) \end{vmatrix} > 0 \tag{2.10}$$

for all $w_1 < \cdots < w_n$ and all $n \leqq m$.

PROOF. The proof proceeds by mathematical induction. Clearly the lemma holds for $m = 1$. Suppose it holds up to $m - 1$, and suppose $w_1 < w_2 < \cdots < w_n$ are given. Let $q_i(w) = p_i(w)/p_1(w)$. Then

$$\begin{aligned} \operatorname{sgn} \begin{vmatrix} p_1(w_1) & \cdots & p_1(w_1) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ p_1(w_1) & \cdots & p_1(w_1) \end{vmatrix} &= \operatorname{sgn} \begin{vmatrix} 1 & \cdots & 1 \\ q_2(w_1) & \cdots & q_2(w_n) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ q_n(w_1) & \cdots & q_n(w_n) \end{vmatrix} \\ &= \operatorname{sgn} \begin{vmatrix} 1 & 0 & \cdots & 0 \\ q_2(w_1) & q_2^1(u_2) & \cdots & q_2^1(u_n) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ q_n(w_1) & q_n^1(u_2) & \cdots & q_n^1(u_n) \end{vmatrix}, \end{aligned} \tag{2.11}$$

where $w_1 < u_2 < w_2 < u_3 < \cdots < u_n < w_n$. But

$$\frac{d^j}{dw^j}\frac{f(w)}{p_1(w)} = \frac{1}{p_1(w)}\frac{d^j}{dw^j}f(w) + \sum_{k=1}^{j-1} a_k(w)\frac{d^k}{dw^k}f(w)$$

where the $a_k(w)$'s do not depend on f. Therefore, for all w

$$\text{(2.12)}\quad \operatorname{sgn}\begin{vmatrix} p_1(w) & p_1^1(w) & \cdots & p_1^{n-1}(w) \\ p_2(w) & p_2^1(w) & \cdots & p_2^{n-1}(w) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ p_n(w) & p_n^1(w) & \cdots & p_n^{n-1}(w) \end{vmatrix} = \operatorname{sgn}\begin{vmatrix} p_1(w) & 0 & \cdots & 0 \\ p_2(w) & q_2^1(w) & \cdots & q_2^{n-1}(w) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ p_n(w) & q_n^1(w) & \cdots & q_n^{n-1}(w) \end{vmatrix}.$$

Since the first determinant in (2.12) is positive by assumption, for all w and $n \leqq m$

$$\begin{vmatrix} q_2^1(w) & \cdots & q_2^{n-1}(w) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ q_n^1(w) & \cdots & q_n^{n-1}(w) \end{vmatrix} > 0$$

By the induction assumption this implies

$$\begin{vmatrix} q_2^1(u_2) & \cdots & q_2^1(u_n) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ q_n^1(u_2) & \cdots & q_n^1(u_n) \end{vmatrix} > 0$$

for all $u_2 < \cdots < u_n$ and all $n \leqq m$. But this determinant equals the last determinant in (2.11) so (2.10) follows.

LEMMA 3. *If p is strictly Pólya Type ∞ and all the derivatives involved exist and are continuous, then*

$$\operatorname{sgn}\det \| p(x_j, \omega_i) \| = \operatorname{sgn} q_n(\xi, \omega_{n+1}, \omega_n, \cdots, \omega_1),$$

$x_1 < x_2 < \cdots < x_{n+1}$, for some appropriate ξ satisfying $x_1 < \xi < x_{n+1}$, *where*

$$q_1(\xi, \omega, \omega_1) = \frac{d}{dx}\left\{\frac{p(\xi, \omega)}{p(\xi, \omega_1)}\right\}, \quad q_2(\xi, \omega, \omega_2, \omega_1) = \frac{d}{dx}\left\{\frac{\dfrac{d}{dx}\dfrac{p(\xi, \omega)}{p(\xi, \omega_1)}}{\dfrac{d}{dx}\dfrac{p(\xi, \omega_2)}{p(\xi, \omega_1)}}\right\},$$

and

$$q_k(\xi, \omega, \omega_k, \omega_{k-1}, \cdots, \omega_1) = \frac{d}{dx}\left\{\frac{q_{k-1}(\xi, \omega, \omega_{k-1}, \cdots, \omega_1)}{q_{k-1}(\xi, \omega_k, \omega_{k-1}, \cdots, \omega_1)}\right\},$$

where $\omega_1 < \omega_2 < \cdots < \omega_n$ *but* ω *and* ω_{n+1} *are allowed to occur anywhere. (The notation means that the derivatives are taken with respect to* x *and evaluated at* $x = \xi$.)

PROOF. The proof proceeds by induction. Let $p(x_j, \omega_i) = p_i(x_j)$. For $n = 1$

$$\operatorname{sgn}\begin{vmatrix} p_1(x_1) & p_1(x_2) \\ p_2(x_1) & p_2(x_2) \end{vmatrix} = \operatorname{sgn}\begin{vmatrix} 1 & 1 \\ \dfrac{p_2(x_1)}{p_1(x_1)} & \dfrac{p_2(x_2)}{p_1(x_2)} \end{vmatrix} = \operatorname{sgn}\begin{vmatrix} 1 & 0 \\ \dfrac{p_2(x_1)}{p_1(x_1)} & \dfrac{d}{dx}\dfrac{p_2(\xi)}{p_1(\xi)} \end{vmatrix}$$

by Lemma 1. Assume the theorem is true for $n - 1$. Then

$$\operatorname{sgn}\begin{vmatrix} p_1(x_1) & \cdots & p_1(x_{n+1}) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ p_{n+1}(x_1) & \cdots & p_{n+1}(x_{n+1}) \end{vmatrix} = \operatorname{sgn}\begin{vmatrix} 1 & 0 & & 0 \\ \dfrac{p_2(x_1)}{p_1(x_1)} & q_1(\xi_2^1, \omega_2, \omega_1) & \cdots & q_1(\xi_{n+1}^1, \omega_2, \omega_1) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \dfrac{p_{n+1}(x_1)}{p_1(x_1)} & q_1(\xi_2^1, \omega_{n+1}, \omega_1) & \cdots & q_1(\xi_{n+1}^1, \omega_{n+1}, \omega_1) \end{vmatrix}$$

$$= \operatorname{sgn}\begin{vmatrix} q_1(\xi_2^1, \omega_2, \omega_1) & \cdots & q_1(\xi_{n+1}^1, \omega_2, \omega_1) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ q_1(\xi_2^1, \omega_{n+1}, \omega_1) & \cdots & q_1(\xi_{n+1}^1, \omega_{n+1}, \omega_1) \end{vmatrix},$$

where $x_{i-1} < \xi_i^1 < x_i$, $i = 2, \cdots, n+1$. By the induction hypothesis

$$q_1(\xi_j^1, \omega_2, \omega_1) > 0 \text{ for } j = 2, \cdots, n+1.$$

Therefore dividing each row by the first and applying Lemma 1,

$$\operatorname{sgn}\begin{vmatrix} q_1(\xi_2^1, \omega_2, \omega_1) & \cdots & q_1(\xi_{n+1}^1, \omega_2, \omega_1) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ q_1(\xi_2^1, \omega_{n+1}, \omega_1) & \cdots & q_1(\xi_{n+1}^1, \omega_{n+1}, \omega_1) \end{vmatrix}$$

$$= \operatorname{sgn} \begin{vmatrix} q_2(\xi_3'', \omega_3, \omega_2, \omega_1) & \cdots & q_2(\xi_{n+1}'', \omega_3, \omega_2, \omega_1) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ q_2(\xi_3'', \omega_{n+1}, \omega_2, \omega_1) & \cdots & q_2(\xi_{n+1}'', \omega_{n+1}, \omega_2, \omega_1) \end{vmatrix},$$

where $\xi_{i-1}^1 < \xi_i'' < \xi_i^1$, $i = 3, \cdots, n+1$. Continuing in this manner and at each step using the fact that the first row is positive by the induction assumption the sequence terminates with the expression $q_n(\xi, \omega_{n+1}, \omega_n, \cdots, \omega_1)$, $x_1 < \xi < x_{n+1}$, which has the same sign as the original determinant.

ξ is a real number which occurs between x_1 and x_{n+1} and depends on $x_1, x_2, \cdots, x_{n+1}$. Det $\| p(x_j, \omega_i) \|$ will always have the same sign, regardless of the values of the x_j's and ω_i's, just so long as the same order relation exists between them. By the continuity assumptions in Lemma 3, ξ can be found to take on any arbitrary real value by varying $x_1, x_2, \cdots, x_{n+1}$. Hence we have the following lemma.

LEMMA 3a. *If the conditions of Lemma* 3 *are satisfied, then*

$$\operatorname{sgn} \det \| p(x_j, \omega_i) \| = \operatorname{sgn} q_n(x, \omega_{n+1}, \omega_n, \cdots, \omega_1),$$

where $x_1 < x_2 < \cdots < x_{n+1}$, $\omega_1 < \omega_2 < \cdots < \omega_n$, *and where* ω_{n+1} *is allowed to occur anywhere, and* x *is any real number.*

The proof of the converse to Theorem 2 follows readily from Lemma 2. In fact, strict inequality in (2.6) for every $n \leqq m$ implies that p is strictly Pólya Type m by Lemma 2 with $p_i(w) = p(x_i, \omega)$, ω playing the role of w in Lemma 2. Strict inequality in (2.7) implies strict inequality in (2.6) by Lemma 2 with $p_i(w) = (\partial^{i-1}/\partial\omega^{i-1})\, p(x, \omega)$, ω being fixed and x playing the role of w in Lemma 2.

The converse statement in Theorem 2 involves strict inequality in (2.6) and (2.7). What can be said if just (2.6) and (2.7) hold? A positive result can be achieved if the following slight condition holds. If relations (2.6) and (2.7) are valid and if for every ω and $n \leqq m$ they hold with strict inequality for some $x_1 < x_2 < \cdots < x_n$, which may depend upon ω, then we can generally still prove that p is Pólya Type m by use of the following device. Let

$$p_\sigma(x, \omega) = \frac{\int_{-\infty}^{\infty} \phi(x - u, \sigma)\, p(u, \omega)\, d\mu(u)}{\int_{-\infty}^{\infty} \phi(x - u, \sigma)\, d\mu(u)}, \tag{2.13}$$

where ϕ is the normal density function with mean 0 and variance σ. The measure μ is chosen so that the integrals exist and are positive with μ possessing positive

measure everywhere. As $\sigma \to 0$, $p_\sigma(x, \omega) \to p(x, \omega)$ uniformly in any finite interval and

$$(2.14) \qquad \frac{\partial}{\partial \omega} p_\sigma(x, \omega) = \frac{\int_{-\infty}^{\infty} \phi(x - u, \sigma) \frac{\partial}{\partial \omega} p(u, \omega)\, d\mu(u)}{\int_{-\infty}^{\infty} \phi(x - u, \sigma)\, d\mu(u)} \to \frac{\partial}{\partial \omega} p(x, \omega),$$

where we have assumed that (2.14) is valid, i.e., the integral can be differentiated inside the integral sign. But[2]

$$\begin{vmatrix} p_\sigma(x_1, \omega) & \frac{\partial}{\partial \omega} p_\sigma(x_1, \omega) & \cdots & \frac{\partial^{n-1}}{\partial \omega^{n-1}} p_\sigma(x_1, \omega) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ p_\sigma(x_n, \omega) & \frac{\partial}{\partial \omega} p_\sigma(x_n, \omega) & \cdots & \frac{\partial^{n-1}}{\partial \omega^{n-1}} p_\sigma(x_n, \omega) \end{vmatrix}$$

$$= \frac{1}{\left(\int_{-\infty}^{\infty} \phi(x - u, \sigma)\, d\mu(u)\right)^n} \int \cdots \int_{u_1 < \cdots < u_n} \begin{vmatrix} \phi(u_1 - x_1, \sigma) & \cdots & \phi(u_1 - x_n, \sigma) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \phi(u_n - x_1, \sigma) & \cdots & \phi(u_n - x_n, \sigma) \end{vmatrix}$$

$$\cdot \begin{vmatrix} p(u_1, \omega) & \frac{\partial}{\partial \omega} p(u_1, \omega) & \cdots & \frac{\partial^{n-1}}{\partial \omega^{n-1}} p(u_1, \omega) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ p(u_n, \omega) & \frac{\partial}{\partial \omega} p(u_n, \omega) & \cdots & \frac{\partial^{n-1}}{\partial \omega^{n-1}} p(u_n, \omega) \end{vmatrix} d\mu(u_1) \cdots d\mu(u_n).$$

Since ϕ is strictly Pólya Type ∞ for each σ, the first determinant in the integrand is always positive. By assumption the second determinant is not identically zero. Therefore p_σ satisfies the determinant criterion for strictly Pólya Type m densities by Theorem 2, and p is Pólya Type m since $\det \| p(x_1, w_j) \| = \lim_{\sigma \to 0} \det \| p_\sigma(x_i, \omega_j) \|$.

[2] See G. Pólya and G. Szego, *Aufgaben und Lehrsatze aus der Analysis*, Vol. 1, p. 48, Problem 68.

This completes the various characterizations of Pólya Type distributions that will be given here. The remaining theorems and corollaries summarize the main properties of Pólya Type distributions. Theorem 3 and its corollaries are crucial to the decision theory in Parts II and III.

Sec. 3. Basic oscillation theorem for P. T. distribution. The following definition is needed to make the concepts in the theorem precise.

Def. 2: The number of sign changes $V(h)$ of a function $h(\omega)$ is taken to be $\sup_{\omega_1,\cdots\omega_m} N(h(\omega_i))$, where $N(h(\omega_i))$ is the number of changes of sign of the sequence $h(\omega_1), h(\omega_2), \cdots, h(\omega_n)$, $\omega_i < \omega_{i+1}$. A point ω_0 is called a change point for $h(\omega)$ if $h(\omega)\ h(\omega') \leqq 0$ whenever $\omega \leqq \omega_0 \leqq \omega'$ with $\omega \neq \omega'$ (ω, ω' essentially near ω_0) and definite inequality occurs for some specific choice of ω and ω' *or* $h(\omega_0)\ h(\omega)\ h(\omega') \leqq 0$ for $\omega < \omega_0 < \omega'$.

THEOREM 3. *Let p be strictly Pólya Type ∞ and assume that p can be differentiated n times with respect to x for all ω. Let F be a measure on the real line, and let h be a function of ω which changes sign n times. If*

$$g(x) = \int p(x, \omega)h(\omega)\, dF(\omega)$$

can be differentiated n times with respect to x inside the integral sign, then g changes sign at most n times and has at most n zeros counting multiplicities or is identically zero. The function g is identically zero if and only if the spectrum of F is contained in the set of zeros of h.

PROOF. Let $\omega_1, \omega_2, \cdots, \omega_n$ be the change points of h. $\omega_1 < \omega_2 < \cdots < \omega_n$. Form

$$\frac{d}{dx}\left\{\frac{g(x)}{p(x, \omega_1)}\right\} = \int \frac{d}{dx}\frac{p(x, \omega)}{p(x, \omega_1)}\, h(\omega)\, dF(\omega)$$

$$\vdots \qquad\qquad \vdots$$

(2.15)

$$q_n^*(x, \omega_n, \omega_{n-1}, \cdots, \omega_1) = \int q_n(x, \omega, \omega_n, \cdots, \omega_1)h(\omega)\, dF(\omega).$$

The function $q_n^*(x, \omega_n, \cdots, \omega_1)$ is the function $q_n(x, \omega, \omega_n, \cdots, \omega_1)$ with $p(x, \omega)$ replaced by $g(x)$. All the above integrands are well-defined since Lemma 3 can be applied.

Suppose for definiteness $h(\omega) > 0$ for $\omega < \omega_1$ and n even. Then $\det \| p(x_j, \omega_i) \|$, $i, j = 1, 2, \cdots, n+1$, with $\omega_{n+1} = \omega$, and $x_1 < x_2 < \cdots < x_{n+1}$ has the same sign as the determinant obtained from the above with first and last rows interchanged. This last determinant is positive as p is assumed to be Pólya Type ∞. Hence, by Lemma 3 $q_n(x, \omega, \omega_n, \cdots, \omega_1) > 0$ for $\omega < \omega_1$ so the integrand in (2.15) is positive. For $\omega_1 < \omega < \omega_2$ with $\omega = \omega_{n+1}$ the original $\det \| p(x_j, \omega_i) \|$ has the opposite sign of the determinant which has the last row inserted between the first and second rows in this determinant. This second determinant is positive

so det $\| p(x_j, \omega_i) \| < 0$. Applying Lemma 3 again we see that the integrand is positive for $\omega_1 < \omega < \omega_2$. Repeating this line of argument we find that $h(\omega)$ and $q_n(x, \omega, \omega_n, \cdots, \omega_1)$ have the same sign so the integrand in (2.15) is always positive. Therefore $q_n^*(x, \omega_n, \omega_{n-1}, \cdots, \omega_1) > 0$ for all x. Now, $q_{n-1}(x, \omega_{n-1}, \omega_{n-2}, \cdots, \omega_1)$ is likewise positive for all x, by Lemma 3. Since $q_n^*(x, \omega_n, \cdots, \omega_1) > 0$ and $q_{n-1}(x, \omega_{n-1}, \cdots, \omega_1) > 0$ for all x, from the definition of $q_{n-1}^*(x, \omega_{n-1}, \cdots, \omega_1)$ we deduce that this function changes sign at most once and has at most one zero. Similarly, since $q_{n-2}(x, \omega_{n-2}, \cdots, \omega_1) > 0$ for all x, this implies that $q_{n-2}^*(x, \omega_{n-2}, \cdots, \omega_1)$ changes sign at most twice and has at most two zeros counting multiplicities. The end of this sequence of implications is that $g(x)$ changes sign at most n times and has at most n zeros counting multiplicities.

Suppose $h(\omega) > 0$ for $\omega < \omega_1$, but n is odd. Then reasoning analogous to that used in the even case shows that the integrand in (2.15) is always negative. Thus $q_n^*(x, \omega, \omega_n, \cdots, \omega_1) < 0$ for all x, and this implies the desired conclusion. A similar argument proves the result when $h(\omega) < 0$ for $\omega < \omega_1$.

By following the sequence of implications in reverse order it can be checked that if g changes sign n times, then it changes sign in the same order as $h(\omega)$. This gives us Corollary 2.

COROLLARY 2. *If the number of sign changes of g is $n = V(h)$, then g and h change signs in the same order.*

COROLLARY 3. *If p is Pólya Type ∞ but not strictly so, the results of Theorem 3 still hold if for any n and any prescribed $\omega_1 < \cdots < \omega_n$ there exists a set of $x_1 < \cdots < x_n$ (which may depend on $\omega_1, \cdots, \omega_n$) such that* det $\| p(x_i, \omega_j) \| > 0$.

This can be established by approximating $p(x, \omega)$ by $p_\epsilon(x, \omega)$ as in (2.13).

The condition that p be strictly Pólya Type ∞ can be weakened also in another manner different from Corollary 3. The results of the theorem still hold if p is strictly Pólya Type $n + 1$, one more than the number of sign changes of h.

Completely analogous results can also be proved about the function

$$g(\omega) = \int p(x, \omega)h(x)\, d\mu(x).$$

Sec. 4. Addition theorem for P. T. distribution. The following two theorems present results which are interesting per se but which will not be of any use in the subsequent sections. These theorems illustrate some other nice smoothening properties possessed by Pólya frequency functions.

THEOREM 4. *Let X and Y be independent real random variables having continuously differentiable densities f and g, and let $f(x - \omega) = f^*(x, \omega)$ be strictly Pólya Type ∞. If g has k modes, then the density of $z = x + y$,*

$$h(z) = \int_{-\infty}^{\infty} f(t)g(z - t)\, dt,$$

has at most k modes. Furthermore, for any constant c, $h - c$ changes sign no more often than $g - c$.

To expedite the discussion we assume that differentiation can be performed underneath the integral.

PROOF.

$$\frac{d}{dz}h(z) = \int_{-\infty}^{\infty} f(t)g'(x-t)\,dt = \int_{-\infty}^{\infty} f(x-y)g'(y)\,dy.$$

Since the number of modes of a density is bounded above by the number of changes of sign of its derivative, the first conclusion follows from Theorem 3.

The second conclusion also follows from Theorem 3 since

$$h(z) - c = \int_{-\infty}^{\infty} f(t)[g(z-t) - c]\,dt = \int_{-\infty}^{\infty} f(x-y)[g(y) - c]\,dy.$$

Z. W. Birnbaum calls a real random variable X less peaked than another real random variable Y if $\Pr\{|X| \leqq u\} \leqq \Pr\{|Y| \leqq u\}$ for all $u > 0$. He proves that if X is less peaked than Y and Z is independent of X and Y and has a symmetric unimodal density, then $X + Z$ is less peaked than $X + Y$ (Ref. [3]). We can generalize this definition and with the aid of Theorem 3 generalize the result.

Def. 3. A real random variable X is less peaked of order n than another real random variable Y if $q(u) = \Pr\{|X| \leqq u\} - \Pr\{|Y| \leqq u\}$ changes sign n times and is nonpositive for sufficiently large u or else changes sign less than n times.

Birnbaum's definition of less peaked corresponds to less peaked of order 0.

THEOREM 5. *Let X be less peaked of order n than Y. If Z is independent of X and Y and has a density h which is symmetric and is such that $h(z-\omega) = h^*(z,\omega)$ is strictly Pólya Type ∞, then $X + Z$ is less peaked of order n than $Y + Z$.*

PROOF. If F and G are the cdf's of X and Y respectively, then

$$\Pr\{|X+Z| \leqq u\} - \Pr\{|Y+Z| \leqq u\} = \int_{-\infty}^{\infty} [F(s) - F(-s) - G(s) + G(-s)]h(u-s)\,ds.$$

The first factor in the integrand is an odd function of s which changes sign at most n times for positive s and hence at most $2n + 1$ times altogether. The second factor is a symmetric, strict Pólya Type ∞ density function. By Theorem 3 the integral is an antisymmetric function of u which changes sign at most $2n + 1$ times and hence at most n times for positive u. Furthermore, if it changes sign n times for positive u, it changes $2n + 1$ times altogether and must therefore have the same sign for very large u that $F(s) - F(-s) - G(s) + G(-s)$ does for very large s.

Part II. Application of Pólya Type Distributions to Classical Results of the Neyman Pearson Variety.

Sec. 1. Preliminaries. A number of classical results can be derived when the underlying distribution is Pólya Type. These results concern Type A regions,

uniformly most powerful tests, unbiased tests, the likelihood ratio test, etc. They unify and strengthen essentially all previously known results. A great deal of the literature on the theory of testing statistical hypotheses deals with special cases [4], whereas this approach is of a more general nature and yields much stronger results and at the same time constructive methods in determining the specialized tests.

The general situation we are dealing with is that of testing a null hypothesis against its alternative hypothesis, i.e., a 2-action problem. The parameter space Ω is the real line. There exist two measurable loss functions L_1 and L_2 on Ω where $L_i(\omega)$ is the loss incurred if action i is taken and ω is the true parameter point. The set in which $L_1(\omega) < L_2(\omega)$ is the set in which action 1 is preferred when ω is the true state of nature, and the set in which $L_2(\omega) < L_1(\omega)$ is the set in which action 2 is preferred. The two actions are indifferent at all other points. We shall assume that $L_1 - L_2 = h$ changes sign exactly n times where n will vary according to the problem we are considering but will remain constant within each problem. The points where $L_1 - L_2$ changes sign are assumed isolated and are ω_1^0, ω_2^0, $\cdots$, ω_n^0. For the sake of definiteness we shall assume that $L_1(\omega) - L_2(\omega)$ is positive for $\omega < \omega_1^0$. Two successive ω_i^0's may be equal but not more than two. In fact, if $\omega_i^0 = \omega_{i+1}^0$, then $[L_1(\omega) - L_2(\omega)][L_1(\omega') - L_2(\omega')] > 0$ for $\omega < \omega_i^0 < \omega'$ (ω, ω' near ω_i^0) and $[L_1(\omega) - L_2(\omega)] \cdot [L_1(\omega_i^0) - L_2(\omega_i^0)] < 0$ for the same choice of ω. This corresponds to the case where one action is preferred in a neighborhood of ω_i^0 except for $\omega = \omega_i^0$ where the other action is preferred.

Let ϕ be a randomized decision procedure. ϕ is a measurable function on the real line, and $\phi(x)$ is the probability of taking action 2 (accepting the alternative hypothesis) if x is the observed value of the real random variable X. (x is usually a sufficient statistic based on several observations.) Consider decision procedures ϕ of the form

$$\phi(x) = \begin{cases} 1 \text{ for } x_{2i} < x < x_{2i+1}, & i = 0, 1, \cdots, \left[\frac{n}{2}\right] \\ \lambda_j \text{ for } x = x_j, 0 \leqq \lambda_j \leqq 1, & j = 1, 2, \cdots, n \\ 0 \text{ elsewhere} \end{cases}$$

$[a]$ denotes the greatest integer $\leqq a$. $x_0 = -\infty$. All randomized decision procedures of this form will be said to belong to the class $\mathfrak{M}_n$ of monotone procedures. If the x_j's are all distinct, then action 2 is preferred in n intervals, action 1 in n or $n - 1$, and at n points there is possible randomization. Strategies with fewer intervals but essentially the same form also belong to $\mathfrak{M}_n$; this corresponds to the case where the x_j's are not all distinct.

The following theorem and lemma will be used in the subsequent discussion. For proofs and greater detail the reader is referred to [1]. It should be remarked that the proofs of Theorem 6 and Lemma 4 can be based essentially on Theorem 3.

THEOREM 6. *If $p(x, \omega)$ belongs strictly to $\mathcal{P}_{n+1}$, then for any randomized decision*

procedure ϕ not in $\mathfrak{M}_n$ there exists a unique ϕ^0 in $\mathfrak{M}_n$ such that $\rho(\omega, \phi^0) \leqq \rho(\omega, \phi)$ with inequality everywhere except for $\omega = \omega_0^1, \omega_2^0, \cdots, \omega_n^0$. ρ is given by $\rho(\omega, \phi) = \int[(1 - \phi(x))L_1(\omega) + \phi(x)\ L_2(\omega)]\ p(x, \omega)\ d\mu(x)$. Moreover, the set $\mathfrak{M}_n$ constitutes a minimal complete class of strategies.

If the underlying distribution $p(x, \omega)$ does not strictly belong to $\mathcal{P}_{n+1}$, then the strategies of $\mathfrak{M}_n$ still constitute a complete class but the uniqueness and minimality conclusion of Theorem 6 is not valid in general. However, by a general device of approximating non strict Pólya Type distributions by strict Pólya Type (see Part I), many of the foregoing results can be extended. This shall be left as an exercise for the reader.

LEMMA 4. *If ϕ^1 and ϕ^2 are two strategies in $\mathfrak{M}_n$ and p is strictly Pólya Type $n + 1$, then*

$$\int [\phi^1(x) - \phi^2(x)]p(x, \omega)\ d\mu(x)$$

has less than n zeros counting multiplicities.

In the future, when we say assume strictly Pólya Type n, we mean that the underlying distribution belongs strictly to $\mathcal{P}_n$.

Sec. 2. Uniformly most powerful one-sided tests. The case of uniformly most powerful tests for the classical exponential family of distributions and other specific examples was treated in Lehman's notes [4]. This represents a slight extension to the situation of Pólya Type distributions.

Assume strictly $\mathcal{P}_2$. For a one-sided testing problem, a uniformly most powerful level α test exists.

A one-sided testing problem occurs when

$$L_1(\omega) = \begin{cases} 1 & \omega < \omega_1 \\ 0 & \omega \geqq \omega_1 \end{cases} \quad \text{and} \quad L_2(\omega) = \begin{cases} 0 & \omega < \omega_1 \\ 1 & \omega \geqq \omega_1 \end{cases}$$

for some ω_1. Then $\rho(\omega, \phi) = \int \phi(x)p(x, \omega)\ d\mu(x)$ for $\omega \geqq \omega_1$ and $\rho(\omega, \phi) = \int (1 - \phi(x))\ p(x, \omega)\ d\mu(x)$ for $\omega < \omega_1$. Consider the function $f_\phi(\omega) = \int \phi(x)p(x, \omega)\ d\mu(x)$ where $\phi \varepsilon\ \mathfrak{M}_1$. $f_\phi(\omega) - c = \int (\phi(x) - c)\ p(x, \omega)\ d\mu(x)$ where c is an arbitrary positive constant. Since $\phi\ \varepsilon\ \mathfrak{M}_1$, $\phi - c$ changes sign at most once and in the direction from $+$ to $-$ if at all. Therefore by Theorem 3 and Corollary 2, $f_\phi - c$ changes sign at most once and in the same direction if at all. This implies that f_ϕ is a monotone decreasing function of ω. Consider that unique monotone test ϕ^* (unique $[\mu]$) for which $f_{\phi^*}(\omega_1) = \int \phi^*(x)p(x, \omega_1)\ d\mu(x) = \alpha$. For any other level α monotone test ϕ_1 the corresponding f_{ϕ_1} is uniformly smaller than f_{ϕ^*} by Lemma 4 so that ϕ^* is best among the monotone tests. Now consider any non-monotone level α test ϕ. By Theorem 6 there is a unique monotone test ϕ_2 which is better than ϕ except at ω_1, where equality holds. But since $f_{\phi^*} \geqq f_{\phi_2}$, ϕ^* also improves on ϕ.

uniformly most powerful tests, unbiased tests, the likelihood ratio test, etc. They unify and strengthen essentially all previously known results. A great deal of the literature on the theory of testing statistical hypotheses deals with special cases [4], whereas this approach is of a more general nature and yields much stronger results and at the same time constructive methods in determining the specialized tests.

The general situation we are dealing with is that of testing a null hypothesis against its alternative hypothesis, i.e., a 2-action problem. The parameter space Ω is the real line. There exist two measurable loss functions L_1 and L_2 on Ω where $L_i(\omega)$ is the loss incurred if action i is taken and ω is the true parameter point. The set in which $L_1(\omega) < L_2(\omega)$ is the set in which action 1 is preferred when ω is the true state of nature, and the set in which $L_2(\omega) < L_1(\omega)$ is the set in which action 2 is preferred. The two actions are indifferent at all other points. We shall assume that $L_1 - L_2 = h$ changes sign exactly n times where n will vary according to the problem we are considering but will remain constant within each problem. The points where $L_1 - L_2$ changes sign are assumed isolated and are ω_1^0, ω_2^0, $\cdots$, ω_n^0. For the sake of definiteness we shall assume that $L_1(\omega) - L_2(\omega)$ is positive for $\omega < \omega_1^0$. Two successive ω_i^0's may be equal but not more than two. In fact, if $\omega_i^0 = \omega_{i+1}^0$, then $[L_1(\omega) - L_2(\omega)][L_1(\omega') - L_2(\omega')] > 0$ for $\omega < \omega_i^0 < \omega'$ (ω, ω' near ω_i^0) and $[L_1(\omega) - L_2(\omega)] \cdot [L_1(\omega_i^0) - L_2(\omega_i^0)] < 0$ for the same choice of ω. This corresponds to the case where one action is preferred in a neighborhood of ω_i^0 except for $\omega = \omega_i^0$ where the other action is preferred.

Let ϕ be a randomized decision procedure. ϕ is a measurable function on the real line, and $\phi(x)$ is the probability of taking action 2 (accepting the alternative hypothesis) if x is the observed value of the real random variable X. (x is usually a sufficient statistic based on several observations.) Consider decision procedures ϕ of the form

$$\phi(x) = \begin{cases} 1 \text{ for } x_{2i} < x < x_{2i+1}, & i = 0, 1, \cdots, \left[\frac{n}{2}\right] \\ \lambda_j \text{ for } x = x_j, 0 \leqq \lambda_j \leqq 1, & j = 1, 2, \cdots, n \\ 0 \text{ elsewhere} \end{cases}$$

$[a]$ denotes the greatest integer $\leqq a$. $x_0 = -\infty$. All randomized decision procedures of this form will be said to belong to the class $\mathfrak{M}_n$ of monotone procedures. If the x_j's are all distinct, then action 2 is preferred in n intervals, action 1 in n or $n - 1$, and at n points there is possible randomization. Strategies with fewer intervals but essentially the same form also belong to $\mathfrak{M}_n$; this corresponds to the case where the x_j's are not all distinct.

The following theorem and lemma will be used in the subsequent discussion. For proofs and greater detail the reader is referred to [1]. It should be remarked that the proofs of Theorem 6 and Lemma 4 can be based essentially on Theorem 3.

THEOREM 6. *If $p(x, \omega)$ belongs strictly to $\mathcal{P}_{n+1}$, then for any randomized decision*

procedure ϕ not in $\mathfrak{M}_n$ there exists a unique ϕ^0 in $\mathfrak{M}_n$ such that $\rho(\omega, \phi^0) \leqq \rho(\omega, \phi)$ with inequality everywhere except for $\omega = \omega_0^1, \omega_2^0, \cdots, \omega_n^0$. ρ is given by $\rho(\omega, \phi) = \int[(1 - \phi(x))L_1(\omega) + \phi(x)\, L_2(\omega)]\, p(x, \omega)\, d\mu(x)$. Moreover, the set $\mathfrak{M}_n$ constitutes a minimal complete class of strategies.

If the underlying distribution $p(x, \omega)$ does not strictly belong to $\mathcal{P}_{n+1}$, then the strategies of $\mathfrak{M}_n$ still constitute a complete class but the uniqueness and minimality conclusion of Theorem 6 is not valid in general. However, by a general device of approximating non strict Pólya Type distributions by strict Pólya Type (see Part I), many of the foregoing results can be extended. This shall be left as an exercise for the reader.

LEMMA 4. *If ϕ^1 and ϕ^2 are two strategies in $\mathfrak{M}_n$ and p is strictly Pólya Type $n + 1$, then*

$$\int [\phi^1(x) - \phi^2(x)]p(x, \omega)\, d\mu(x)$$

has less than n zeros counting multiplicities.

In the future, when we say assume strictly Pólya Type n, we mean that the underlying distribution belongs strictly to $\mathcal{P}_n$.

Sec. 2. Uniformly most powerful one-sided tests. The case of uniformly most powerful tests for the classical exponential family of distributions and other specific examples was treated in Lehman's notes [4]. This represents a slight extension to the situation of Pólya Type distributions.

Assume strictly $\mathcal{P}_2$. For a one-sided testing problem, a uniformly most powerful level α test exists.

A one-sided testing problem occurs when

$$L_1(\omega) = \begin{cases} 1 & \omega < \omega_1 \\ \\ 0 & \omega \geqq \omega_1 \end{cases} \quad \text{and} \quad L_2(\omega) = \begin{cases} 0 & \omega < \omega_1 \\ \\ 1 & \omega \geqq \omega_1 \end{cases}$$

for some ω_1. Then $\rho(\omega, \phi) = \int\phi(x)p(x, \omega)\, d\mu(x)$ for $\omega \geqq \omega_1$ and $\rho(\omega, \phi) = \int (1 - \phi(x))\, p(x, \omega)\, d\mu(x)$ for $\omega < \omega_1$. Consider the function $f_\phi(\omega) = \int\phi(x)p(x, \omega)\, d\mu(x)$ where $\phi\varepsilon\, \mathfrak{M}_1$. $f_\phi(\omega) - c = \int(\phi(x) - c)\, p(x, \omega)\, d\mu(x)$ where c is an arbitrary positive constant. Since $\phi\, \varepsilon\, \mathfrak{M}_1$, $\phi - c$ changes sign at most once and in the direction from $+$ to $-$ if at all. Therefore by Theorem 3 and Corollary 2, $f_\phi - c$ changes sign at most once and in the same direction if at all. This implies that f_ϕ is a monotone decreasing function of ω. Consider that unique monotone test ϕ^* (unique $[\mu]$) for which $f_{\phi^*}(\omega_1) = \int\phi^*(x)p(x, \omega_1)\, d\mu(x) = \alpha$. For any other level α monotone test ϕ_1 the corresponding f_{ϕ_1} is uniformly smaller than f_{ϕ^*} by Lemma 4 so that ϕ^* is best among the monotone tests. Now consider any non-monotone level α test ϕ. By Theorem 6 there is a unique monotone test ϕ_2 which is better than ϕ except at ω_1, where equality holds. But since $f_{\phi^*} \geqq f_{\phi_2}$, ϕ^* also improves on ϕ.

Sec. 3. Nonexistence of uniformly most powerful two-sided tests. Assume strictly $\mathcal{P}_2$. For the two-sided testing problem uniformly most powerful level α tests do not exist in general. We discuss this now in greater detail.

A two-sided testing problem is determined by

$$L_1(\omega) = \begin{cases} 0 & \omega_1 \leqq \omega \leqq \omega_2 \\ & \\ 1 & \text{elsewhere} \end{cases} \qquad \text{and} \qquad L_2(\omega) = \begin{cases} 1 & \omega_1 \leqq \omega \leqq \omega_2 \\ & \\ 0 & \text{elsewhere} \end{cases}$$

for some $\omega_1 \leqq \omega_2$. By virtue of Theorem 6 we can restrict our consideration exclusively to monotone tests, i.e., tests in $\mathfrak{M}_2$. Let ϕ_3 be a monotone test and $f_{\phi_3}(\omega) = \int \phi_3(x)\, p(x, \omega)\, d\mu(x)$ be the corresponding power function. Consider the one-sided testing problem obtained from the two-sided problem above by removing one tail.

$$L_1^*(\omega) = \begin{cases} 0 & \omega \geqq \omega_1 \\ & \\ 1 & \omega < \omega_1 \end{cases} \qquad \text{and} \qquad L_2^*(\omega) = \begin{cases} 1 & \omega \geqq \omega_1 \\ & \\ 0 & \omega < \omega_1 \end{cases}$$

The existence of a u.m.p. level α test ϕ^* for this problem was shown in section 2. Suppose $\phi_3 \varepsilon\, \mathfrak{M}_1$. Then $f_{\phi^*}(\omega) > f_{\phi_3}(\omega)$ for $\omega < \omega_1$, and ϕ_3 is not u.m.p. for the two-sided test. The strict inequality is assured by Theorem 6. Suppose $\phi_3 \varepsilon\, \mathfrak{M}_1$. Then f_{ϕ_3} is monotone decreasing which means that for $\omega > \omega_2$ the test $\phi \equiv \alpha$ is better.

A word should be said about what happens when p is not strictly Pólya Type 3. When $\omega_1 = \omega_2$ and $P(x, \omega)$ is the rectangular distribution on the interval $[0, \omega]$, a u.m.p. test exists. The acceptance region is $[x', \omega_1]$ where $x' = \alpha\omega_1$. When $P(x, \omega)$ is the rectangular distribution on $[0, \omega]$ but if $\omega_1 < \omega_2$, no u.m.p. test exists; i.e., when the null hypothesis is an interval no u.m.p. test exists. The rectangular distribution on $[0, \omega]$ is Pólya Type 3, but it is not strictly so. It even satisfies the condition that for every $\omega_1 < \omega_2 < \omega_3$ there exists a set $x_1 < x_2 < x_3$ such that $\det \| p(x_i, \omega_j) \| > 0$. Thus the condition of strictness in this result seems very essential.

Sec. 4. Uniformly most powerful unbiased tests.

(a) Assume strictly $\mathcal{P}_2$. For a two-sided testing problem a u.m.p. unbiased test exists. For this special testing problem the result is known for scattered examples.

A test ϕ is unbiased if and only if $f_\phi(\omega) \leqq \alpha$ for $\omega_1 \leqq \omega \leqq \omega_2$ and $f_\phi(\omega) \geqq \alpha$ for $\omega \leqq \omega_1$ and $\omega \geqq \omega_2$. Consider the test $\phi \equiv \boldsymbol{\alpha}$. By Theorem 6 there exists a unique test $\phi^* \varepsilon\, \mathfrak{M}_2$ which uniformly improves in terms of risk over $\phi \equiv \alpha$ except at ω_1 and ω_2. Clearly ϕ^* is unbiased. ϕ^* is determined by x_1^*, x_2^*, λ_1^* and λ_2^* which are the values satisfying $\int \phi^*(x)\, p(x, \omega_i)\, d\mu(x) = \alpha$, $i = 1, 2$, where $\omega_1 < \omega_2$ and x_1^*, x_2^*, λ_1^*, and λ_2^*, determined satisfying $\int \phi^*(x)\, p(x, \omega_1)\, d\mu(x) = \alpha$ and $(d/d\omega) \int \phi^*(x)\, p(x, \omega)\, d\mu(x) \mid_{\omega_1} = 0$ if $\omega_1 = \omega_2$. When $\omega_1 = \omega_2$ the null hypothesis con-

sists of a single point. Lemma 4 shows that fixing the value of f_{ϕ^*} at two points or fixing f_{ϕ^*} and its derivative at one point is sufficient to determine ϕ^* uniquely. Suppose there were an unbiased level α test ϕ for which $f_\phi(\omega) > f_{\phi^*}(\omega)$ for some $\omega < \omega_1$ or $\omega > \omega_2$. Then there would have to be a monotone test ϕ_4 which improves on ϕ except at ω_1 and ω_2. But this contradicts the fact that ϕ^* is the unique monotone test uniformly better than $\phi \equiv \alpha$ except for ω_1 and ω_2. Thus ϕ^* is the u.m.p. unbiased level α test.

(b) Assume strictly $\mathcal{P}_{n+1'}$. For any preference pattern for the two-action testing problem say involving $n + 1$ distinct regions where action 1 is favored, a u.m.p. unbiased test exists.

The above argument generalizes easily to any preference pattern. The unique test $\phi^* \varepsilon \mathfrak{M}_n$ which is uniformly better than $\phi \equiv \alpha$ except at $\omega_1, \omega_2, \cdots, \omega_n$ is the u.m.p. unbiased level α test where ω_i corresponds to the change points of $L_1 - L_2$. ϕ^* is uniquely determined by solving the system of equations $\int \phi^*(x)\, p(x, \omega_i)\, d\mu(x) = \alpha$, $i = 1, 2, \cdots, n$ for $x_1^*, x_2^*, \cdots, x_n^*, \lambda_1^*, \lambda_2^*, \cdots, \lambda_n^*$. For the case where $\omega_i = \omega_{i+1}$ for some i replace the equation $\int \phi_{\alpha^*}(x)\, p(x, \omega_{i+1})\, d\mu(x) = \alpha$ by $(d/d\omega) \int \phi^*(x)\, p(x, \omega)\, d\mu(x) \mid_{\omega_i} = 0$. Lemma 4 shows that the system of equations in this latter case is still sufficient to determine ϕ^* uniquely.

Sec. 5. Generalization of unbiased tests. Assume strictly $\mathcal{P}_{n+1}$, and assume the two-action testing problem under consideration involves $n + 1$ preference regions. Let ϕ^0 be an arbitrary but fixed test. There exists a test u.m.p. with respect to the class Φ_{ϕ^0} of all tests which improve on ϕ^0.

This generalizes the concept of unbiased tests because the class of unbiased tests can be defined as the class of all tests which improve on the test $\phi \equiv \alpha$.

There exists a unique monotone test ϕ^* which improves on ϕ^0. f_{ϕ^*} lies above f_{ϕ^0} in those intervals in which action 2 is preferred to action 1 and it lies below in the other intervals. Some intervals may be degenerate and consist of a single point. Two tests in $\mathfrak{M}_n$ cannot improve on ϕ^0 since they both must have the same power as ϕ^0 at $\omega_1, \omega_2, \cdots, \omega_n$ and this is impossible by Lemma 4. Any nonmonotone test improving on ϕ^0 has a monotone test improving on it by Theorem 6, and this monotone test must be ϕ^*.

Sec. 6. Nature of Type A critical regions. Assume strictly $\mathcal{P}_2$ and assume that for any power function differentiation inside the integral sign with respect to ω is valid. For testing a single point ω_1 against all alternatives a Type A region can be characterized as the union of at most two semi-infinite intervals, i.e., its complement is a single interval.

A Type A critical region is the critical region for any test ϕ which maximizes the curvature $\int \phi(x)\, (\partial^2/\partial\omega^2)\, p(x, \omega) \mid_{\omega_1} d\mu(x)$ subject to the constraints $\int \phi(x)\, p(x, \omega_1)\, d\mu(x) = \alpha$ and $\int \phi(x)\, (\partial/\partial\omega)\, p(x, \omega) \mid_{\omega_1} d\mu(x) = 0$. We know from section 4 that a u.m.p. unbiased level α test exists. In fact it is the unique test $\phi^* \varepsilon \mathfrak{M}_2$ for which $\int \phi^*(x)\, p(x, \omega_1)\, d\mu(x) = 0$ and $(\partial/\partial\omega) \int \phi^*(x)\, p(x, \omega)\, d\mu(x) \mid_{\omega_1} = 0$. By Theorem 6 any nonmonotone test satisfying the constraints has a unique monotone test improving on it and this test must be ϕ^* because of the uniqueness.

Thus ϕ^* defines the Type A critical region, and since ϕ^* belongs to $\mathfrak{M}_2$ the critical region is the union of at most two semi-infinite intervals.

Two remarks should be made. First, all of the known Pólya Type distributions can be differentiated inside the integral sign. Second, for general distributions it is not true that a Type A critical region is the union of two semi-infinite intervals.

Sec. 7. Type A critical regions as a function of the level of significance. Assume strictly $\mathcal{P}_3$ and assume that every power function can be differentiated inside the integral sign with respect to ω. Further, suppose $p(x, \omega) > 0$ for all x and $(\partial/\partial\omega)\ p(x, \omega_1)$ is continuous in x. The assertion is that the complement of the Type A critical region for testing ω_1 against all alternative at level α contains the complement of the Type A critical region for testing ω_1 at level α^1 whenever $\alpha < \alpha^1$. In other words, whenever the hypothesis is rejected for the level of significance α, then it should be rejected also for level α^1 where $\alpha^1 > \alpha$.

This property is not true for Type A regions in general. (See [5].) In order to establish the above result, we need to use the following lemma.

LEMMA 5. *Assume* $\mathcal{P}_2$. *If* $p(x, \omega_1) > 0$ *for all* x *and* $(\partial/\partial\omega)p(x, \omega_1)$ *exists and has at most isolated zeros, then there exists an* x_0 *such that*

$$\frac{\partial}{\partial\omega}\, p(x, \omega_1) \begin{cases} \leqq 0 & x < x_0 \\ \geqq 0 & x > x_0 \end{cases}.$$

PROOF. By Theorem 1

$$\begin{vmatrix} p(x_1, \omega_1) & \dfrac{\partial}{\partial\omega}\, p(x_1, \omega_1) \\ p(x_2, \omega_1) & \dfrac{\partial}{\partial\omega}\, p(x_2, \omega_1) \end{vmatrix} \geqq 0 \tag{3.1}$$

for $x_1 < x_2$. If $(\partial/\partial\omega)\ p(x, \omega_1)$ has no zeros, then the lemma is true with $x_0 = +\infty$. If $(\partial/\partial\omega)\ p(x, \omega_1)$ has zeros, take x_0 to be any one of them. Choose $x_1 = x < x_0 = x_2$. (3.1) reduces to $-p(x_0, \omega_1)\ (\partial/\partial\omega)\ p(x, \omega_1) \geqq 0$. Since $p(x_0, \omega_1) > 0$, we deduce that $(\partial/\partial\omega)\ p(x, \omega_1) \leqq 0$ for $x < x_0$. Now, select $x_1 = x_0 < x = x_2$. (3.1) reduces to $p(x_0, \omega_1)\ (\partial/\partial\omega)\ p(x, \omega_1) \geqq 0$. Since $p(x_0, \omega_1) > 0$, $(\partial/\partial\omega)p(x, \omega_1) \geqq 0$ for $x > x_0$.

From section 6 we know that the Type A critical region for level α is given by a test $\phi_\alpha \varepsilon\, \mathfrak{M}_2$. For simplicity of exposition let us suppose that no randomization is involved and that ϕ_α is defined by the points x_1 and x_2 which must satisfy

$$\int_{x_1}^{x_2} p(x, \omega_1)\, d\mu(x) = 1 - \alpha \text{ and } \int_{x_1}^{x_2} \frac{\partial}{\partial\omega}\, p(x, \omega_1)\, d\mu(x) = 0.$$

We assert that $x_1 < x_0 < x_2$, for otherwise the integrand $\int_{x_1}^{x_2} (\partial/\partial\omega)p(x, \omega_1)\, d\mu(x)$ would be of one sign. (The hypothesis implies that $(\partial/\partial\omega)p(x_1, \omega_1)$ is not identically zero in the interval (x_1, x_2).) Let x_1' and x_2' be the two points defining the

level α' test, where $x_1' < x_0 < x_2'$ with α' near α ($\alpha' > \alpha$) so that x_1' and x_2' are near x_1 and x_2 respectively. Clearly the first constraint prevents the interval (x_1', x_2') from containing the interval (x_1, x_2). Suppose $x_1' \geqq x_1$ and $x_2' > x_2$. Subtracting the second constraint for x_1, x_2 from the second constraint from x_1', x_2' yields

$$\int_{x_2}^{x_2'} \frac{\partial}{\partial \omega} p(x, \omega_1)\, d\mu(x) - \int_{x_1}^{x_1'} \frac{\partial}{\partial \omega} p(x, \omega_1)\, d\mu(x) = 0. \tag{3.2}$$

Between x_2 and x_2' $(\partial/\partial\omega)p(x, \omega_1) > 0$ and between x_1 and x_1' $(\partial/\partial\omega)p(x, \omega_1) < 0$. Hence (3.2) is impossible. A similar argument excludes the case $x_1' < x_1, x_2' \leqq x_2$. Thus $x_1 \geqq x_1'$ and $x_2' \leqq x_2$. The reader can furnish the modifications necessary for the argument when randomization is required at the end points.

Sec. 8. Envelope power function. Assume strictly $\mathcal{P}_3$. For the problem of testing at level α a single point ω_1 against all alternatives the envelope power function $\rho(\omega)$ decreases monotonically away from ω_1 in both directions. Let $\mathfrak{U}_\rho$ be the class of tests ϕ such that if $\rho(\omega') > \rho(\omega'')$ where $\omega_1 \leqq \omega' < \omega''$ or $\omega'' < \omega' \leqq \omega_1$, then $\rho(\omega', \phi) > \rho(\omega'', \phi)$. It will now be established that there exists a test u.m.p. with respect to the class $\mathfrak{U}_\rho$.

Theorem 6 shows that in obtaining the envelope power function the only tests that need be considered are those in $\mathfrak{M}_2$. For $\omega \geqq \omega_1$, $\rho(\omega) = \rho(\omega, \phi^*)$ where ϕ^* is the u.m.p. level α one-sided test of $\omega \leqq \omega_1$ against $\omega > \omega_1$. For $\omega \leqq \omega_1$ $\rho(\omega) = \rho(\omega, \phi^{**})$ where ϕ^{**} is the u.m.p. level α one-sided test of $\omega \geqq \omega_1$ against $\omega < \omega_1$. Lemma 4 shows that no other monotone test can improve on $\rho(\omega, \phi^*)$ for $\omega \geqq \omega_1$ or on $\rho(\omega, \phi^{**})$ for $\omega < \omega_1$. Thus the ρ so defined is actually the envelope power function, and from its definition it is clear that it decreases monotonically away from ω_1 in both directions.

By Section 4 there exists a u.m.p. unbiased level α test. Now any test of $\mathfrak{M}_2$ has the form that the power function can have at most one relative maximum. Indeed, it is sufficient to show that the set of points ω where $\int \phi(x)\, p(x, \omega)\, d\mu(x)$ exceeds any given constant $0 < K < 1$ consists of a single interval. As $\phi(x) - K$ changes sign twice in the order $- + -$, we deduce that the same holds for $\int (\phi(x) - K)\, p(x, \omega)\, d\mu(x) = \int \phi(x)\, p(x, \omega)\, d\mu(x) - K$ from which the conclusion follows. Any unbiased test in $\mathfrak{M}_2$ must therefore also be in $\mathfrak{U}_\rho$, and conversely. The only other possible competing tests that need be worried about are the u.m.p. unbiased level α' tests where $\alpha' < \alpha$. By Section 7 the acceptance region for $\phi_{\alpha'}$ contains the acceptance region for ϕ_α. Thus the probability of rejection at any ω for $\phi_{\alpha'}$ is at least as small as that for ϕ_α. Hence $\rho(\omega, \phi_{\alpha'}) \geqq \rho(\omega, \phi_\alpha)$ for all $\omega \neq \omega_1$.

Sec. 9. The nature of the likelihood ratio test. Assume strictly $\mathcal{P}_2$, and assume $p(x, \omega)$ is a continuous function of x and ω. We prove that the likelihood ratio test is a monotone test.

More explicitly what we mean by the likelihood ratio test being monotone

is the following. Let the null hypothesis be that $\omega \varepsilon \Lambda$ and the alternative hypothesis be that $\omega \varepsilon \Omega$. $\Lambda \cap \Omega = \phi$ and $\Lambda \cup \Omega = R^1$. Suppose Λ is the union of n disjoint intervals some of which may be degenerate, i.e., points. Then

$$I(c) = \left\{ x \;\middle|\; \frac{\sup_{\omega \varepsilon \Lambda} p(x, \omega)}{\sup_{\omega \varepsilon \Omega} p(x, \omega)} \geqq c \right\}$$

is the union of at most n disjoint intervals and hence belongs to $\mathfrak{M}_n$.

Consider a point ω_1 in one of the intervals of Λ. Let $I_{\omega_1} =$

$$\{x \mid p(x, \omega_1) \geqq c \sup_{\omega \varepsilon \Omega} p(x, \omega)\}.$$

That I_{ω_1} depends on c will be understood. $I_{\omega_1} = \bigcap_{\omega \varepsilon \Omega} \{x \mid p(x, \omega_1) \geqq \mathrm{cp}(x, \omega)\}$. Since $p \varepsilon \mathcal{P}_2$, $[p(x, \omega_1)]/[p(x, \omega)]$ is a monotone decreasing (increasing) function of x for $\omega > \omega_1$ ($\omega < \omega_1$). Thus $\{x \mid p(x, \omega_1) \geqq \mathrm{cp}(x, \omega)\}$ is a semi-infinite interval either to the left or right so I_{ω_1} is an interval. If ω_1 is not a degenerate interval of Λ, consider another point ω_2 (for definiteness $\omega_1 < \omega_2$) which is in the same interval of Λ as ω_1. $I_{\omega_2} = \{x \mid p(x, \omega_2) \geqq c \sup_{\omega \varepsilon \Omega} p(x, \omega)\}$ is an interval. Either I_{ω_1} is contained in I_{ω_2}, or I_{ω_2} contains points of I_{ω_1} and points greater than those in I_{ω_1}. It cannot happen that I_{ω_2} contains points less than those in I_{ω_1} without containing all points in I_{ω_1}. Suppose the contrary that this did happen. There then exist two points $x_1 > x_2$ such that $x_2 \varepsilon I_{\omega_2}$, $x_2 \notin I_{\omega_1}$, $x_1 \varepsilon I_{\omega_1}$ and $x_1 \notin I_{\omega_2}$. Since $x_2 \varepsilon I_{\omega_2}$ $[p(x_2, \omega_2)]/[p(x_2, \omega) \geqq c$ for all ω in Ω, and since $x_2 \notin I_{\omega_1}$ there exists a $\omega' \varepsilon \Omega$ such that $[p(x_2, \omega_1)]/[p(x_2, \omega')] < c$. Thus $p(x_2, \omega_2) > p(x_2, \omega_1)$. By a similar argument $p(x_1, \omega_1) > p(x_1, \omega_2)$. This gives $p(x_1, \omega_1)\, p(x_2, \omega_2) - p(x_1, \omega_2)\, p(x_2, \omega_1) > 0$ which is impossible by assumption since $x_1 > x_2$. Thus the assertion is true.

The continuity of $p(x, \omega)$ in both variables simultaneously implies the following continuity property between I_ω and ω. The proof is standard and shall be omitted.

Property 1. Let ω_0 be a fixed point in a nondegenerate interval J of Λ. For every open interval U properly containing I_{ω_0} there exists an $\epsilon > 0$ such that I_ω is contained in the open interval U for all $\omega \varepsilon J$ satisfying $|\omega - \omega_0| < \epsilon$.

Consider any nondegenerate interval $I = (a, b) \varepsilon \Lambda$. It will now be shown that $\bigcup_{\omega \varepsilon I} I_\omega$ is an interval. Suppose to the contrary that there is a point x^* such that $x^* \notin \bigcup_{\omega \varepsilon I} I_\omega$ and there exist I_ω for $\omega \varepsilon I$ above and below x^*. Property 1 and the fact that if $\omega_1 < \omega_2$ I_{ω_2} contains points less than those in I_{ω_1} only if $I_{\omega_1} \subset I_{\omega_2}$ show that the set of $\omega \varepsilon I$ for which I_ω lies above x^* is an open interval if $b \notin I$ and a half-open interval if $b \varepsilon I$. Similarly the set of $\omega \epsilon I$ for which I_ω lies below x^* is an open interval if $a \notin I$, and a half-open interval if $a \varepsilon I$. But if $a, b \notin I$ it is impossible for the interval $I = (a, b)$ to be the union of two disjoint nonempty open intervals. Similar contradictions hold when $a \varepsilon I$, $b \varepsilon I$, and $a, b \varepsilon I$. Therefore $\bigcup_{\omega \varepsilon I} I_\omega$ is an interval. The next interval in Ω to the right of I will produce an interval in the x-space to the right of $\bigcup_{\omega \varepsilon I} I_\omega$ or including $\bigcup_{\omega \varepsilon I} I_\omega$. Repeating this reasoning we see that the proof of our proposition is complete.

It has thus been shown that the likelihood ratio test is a monotone test and hence admissible by virtue of Theorem 6.

Part III. Minimax Strategies for Nature and the Statistician in the General Two-action Problem. A brief summary of the results already obtained in [2] for the one-sided testing problem will first be given for the sake of completeness. The parameter space Ω is an interval (c, d) of the real line. c may be $-\infty$, d may be $+\infty$, and the interval may be open or closed at c or d if either or both are finite. There is a point $\omega_0 \varepsilon (c, d)$ such that action 1 is preferred for $\omega \leqq \omega_0$ and action 2 is preferred for $\omega \geqq \omega_0$. The two loss functions $L_1(\omega)$ and $L_2(\omega)$ are continuous, and $L_1(\omega) = 0$ for $\omega \leqq \omega_0$; > 0 for $\omega > \omega_0$ and $L_2(\omega) = 0$ for $\omega \geqq \omega_0$; > 0 for $\omega < \omega_0$. The risk function ρ is given by

$$(4.1) \qquad \rho(F, \phi) = \iint [L_1(\omega)\phi(x) + L_2(\omega)(1 - \phi(x))]p(x, \omega)\, d\mu(x)\, dF(\omega),$$

where F is the randomized strategy (a priori distribution) for nature and ϕ is the randomized strategy for the statistician. p is assumed to be strictly Pólya Type 2. The following two conditions were required:

Condition 1. If Ω is open at d and a is in the interior of the convex hull of the spectrum of μ, then as $\omega \to d$

$$L_1(\omega) \int_{-\infty}^{a} P(x, \omega)\, d\mu(x) \to 0.$$

Condition 2. If Ω is open at c and b is in the interior of the convex hull of the spectrum of μ, then as $\omega \to c$

$$L_2(\omega) \int_{b}^{\infty} P(x, \omega)\, d\mu(x) \to 0.$$

Under the above assumptions it was shown in [2] that the game $G = (\{F\}, \{\phi\}, \rho)$ has a value and both nature and the statistician have minimax strategies. Moreover the statistician has a monotone minimax strategy and nature has a minimax strategy which concentrates at just two points.

Sec. 1. Minimax theorem in the case when Ω is closed. Our first objective is to present the basic minimax theorem for the general two-action problem in the case where the parameter space Ω is closed. We deal with the situation where there exists $n + 1$ distinct intervals arranged in order in which actions 1 and 2 are successively preferred. When $n = 1$ then our general preference pattern reduces principally to the classical one-sided test of hypothesis. For $n = 2$, we are treating the classical two-sided testing problem. We assume throughout that $L_1(\omega)$ and $L_2(\omega)$ are both continuous. The fundamental preliminary minimax theorem becomes:

THEOREM 7. *If the parameter space Ω is closed, then the game defined by the risk function $\rho(F, \phi)$ is determined (has a value) and the statistician possesses a monotone minimax strategy while nature has a minimax strategy involving at most $n + 1$ points of increase, i.e. the nature's minimax distribution concentrates at most at $n + 1$ points. $n + 1$ is the total number of disconnected preference regions of both actions.*

PROOF. As Ω is closed we know that the space of distributions F over Ω is

compact in the weak* topology with respect to the continuous functions on Ω. This is essentially the Helly selection theorem. Also, the space of strategies ϕ in the two-action problem is also compact in the weak* topology over the integrable functions on X. Obviously, $\rho(F, \theta)$ is linear and continuous with respect to the appropriate weak* topologies and thus optimal strategies F^0 and ϕ^0 exist and the game $\rho(F, \phi)$ has a value (see [6]).

As the class of all monotone strategies constitute a complete class [1], there exists a monotone strategy ϕ^* which improves uniformly on ϕ^0 in terms of risk and hence ϕ^* is minimax. Let

$$T = \{\omega \mid \rho(\omega, \phi^*) = \max_\omega \rho(\omega, \phi^*) = v\},$$

where v is the value of the game. We must distinguish between n odd or even. The analysis will be made for n odd and the details for n even are left for the reader to supply. Suppose for definiteness that the monotonic strategy ϕ^* has the form

$$\phi^*(x) = \begin{cases} 1 & x_{2i-1} < x < x_{2i} \\ & \qquad\qquad i = 1, \cdots, m, \\ 0 & x_{2i} < x < x_{2i+1} \end{cases}$$

with $x_1 = -\infty$ and $x_{2m+1} = +\infty$ and where the x_j are distinct. In other words there are $2m$ disconnected intervals where different preferences of actions 1 and 2 are desired. Of course, m is limited such that $2m \leqq n + 1$ (see Theorem 6).

We now assert that T meets at least $2m$ alternate intervals where actions 1 and 2 are successively preferred. Suppose the contrary: let us consider

$$\int p(x, \omega)[L_1(\omega) - L_2(\omega)]\, dF^0(\omega). \tag{4.2}$$

As $F^0(\omega)$ must concentrate its full measure in T and the only sign changes of $L_1(\omega) - L_2(\omega)$ occur as we pass from one preference region to another, we infer that $[L_1(\omega) - L_2(\omega)]\, dF^0(\omega)$ changes sign less than $2m - 1$ times. Thus, (4.2) by Theorem 3 must change signs fewer than $2m - 1$ times. However, ϕ^* is Bayes against F^0 and must therefore take the values $+1$ or 0 according as (4.2) is negative or positive. Thus ϕ^* cannot have the form as indicated. This contradiction implies the assertion made above about T.

Select $2m$ points T^* in T each belonging to a different preference region such that $L_1(\omega) - L_2(\omega)$ traversing these points changes sign $2m - 1$ times. By Theorem 5 of [1] there exists a distribution F^* which fully concentrates on T^* against which ϕ^* is Bayes. We now show that F^* is minimax. As F^* concentrates in $T^* \subset T$, we get $v = \rho(F^*, \phi^*)$. Using the Baysian nature of ϕ^* for F^*, we obtain

$$v = \rho(F^*, \phi^*) \leqq \rho(F^*, \phi), \qquad \text{for all strategies } \phi.$$

The proof of the theorem is thus complete.

Sec. 2. Two-sided minimax theorem. Our next task is to eliminate the restriction that Ω is closed. For this purpose we need to impose some further conditions on the family of densities $p(x, \omega)$. To expedite and clarify the reasoning, we restrict ourselves to the two-sided problem. Similar analysis would apply to the general two-action problem.

Reviewing the basic assumptions, we have that the parameter space Ω is an interval (c, d) of the real line. c may be $-\infty$, d may be $+\infty$, and the interval may be open or closed at c or d if either or both are finite. There exist two points $\omega_1, \omega_2 \in (c, d)$ such that action 1 is preferred for $\omega \leqq \omega_1$ and $\omega \geqq \omega_2$ and action 2 is preferred for $\omega_1 \leqq \omega \leqq \omega_2$. The two loss functions L_1 and L_2 are continuous, and $L_1(\omega) = 0$ for $\omega \leqq \omega_1, \omega \geqq \omega_2$; > 0 for $\omega_1 < \omega < \omega_2$, and $L_2(\omega) = 0$ for $\omega_1 \leqq \omega \leqq \omega_2$; > 0 for $\omega < \omega_1, \omega > \omega_2$. There is no loss of generality in taking the loss function equal to zero where the action is preferred as differences of the loss functions are the only relevant quantities involved. The risk function is again given by (3.1). This time p is assumed to be strictly Pólya Type 3.

The assertion that will be proven under certain hypothesis of smoothness is that the game $G = (\{F\}, \{\phi\}, \rho)$ has a value and both players have minimax strategies. The statistician has a monotone minimax strategy, and nature has a minimax strategy which concentrates on at most 3 points. To establish this assertion we impose three conditions:

Condition A. If Ω is open at d and a is interior to the convex hull of the spectrum of μ, then as $\omega \to d$

$$L_2(\omega) \int_{-\infty}^{a} P(x, \omega)\, d\mu(x) \to 0.$$

Condition B. If Ω is open at c and b is interior to the convex hull of the spectrum of μ, then as $\omega \to c$

$$L_2(\omega) \int_{b}^{\infty} P(x, \omega)\, d\mu(x) \to 0.$$

These conditions require that if either endpoint is open then as ω tends to this endpoint the mass of probability shifts away from the opposite end of the axis in such a way that the probability at the opposite end of the axis must tend to zero at a faster rate than the loss L blows up. These conditions are similar to those imposed in the one-sided problem.

Condition C. Let $l(c) = \lim_{\omega \to c} L_2(\omega)$ and $l(d) = \lim_{\omega \to d} L_2(\omega)$ (the existence of the limits is postulated).

(i) $l = \min(l(c), l(d)) > \max_{\omega_1 \leqq \omega \leqq \omega_2} L_1(\omega)$

(ii) If Ω is open at c, then $l(c) > \max_{a \leqq \omega \leqq b} L_2(\omega)$ for any closed interval contained in Ω.

(iii) If Ω is open at d, then $l(d) > \max_{a \leqq \omega \leqq b} L_2(\omega)$ for any closed interval contained in Ω.

Condition C has essentially the effect of eliminating the possibility that nature will desire to concentrate at the ends of the parameter space in choosing an

optimal strategy. This condition is fulfilled, for instance, when the losses tend to ∞ at both ends.

Suppose Ω is not a closed interval (the case treated in Theorem 7). Then the result is not immediate because the space $\{F\}$ is not compact. In fact it is no longer true unless for example the three conditions A, B, and C are imposed on the problem; that is, some conditions are necessary. The method of proof consists in considering the sequence of games $G^n = (\{F^n\}, \{\phi\}, \rho)$ where $\Omega^n = [\omega_n', \omega_n'']$ and $\omega_n' \to c$, $\omega_n'' \to d$. That is, we consider a sequence of games defined over closed intervals contained in Ω which in the limit approach Ω. If one end of Ω is closed, say d, then $\omega_n'' \equiv d$ for all n. Each game G^n has a value, and for each game G^n the statistician has a monotone minimax strategy ϕ^n and nature has a strategy (distribution $\bar{F}^n$ which concentrates at at most three points by Theorem 7. The problem is to show that subsequences can be selected from $\{\phi^n\}$ and $\{\bar{F}^n\}$ which converge to strategies yielding the desired properties in the original game G.

Let ν_n be the value of the game G^n. The sequence of ν_n's is bounded away from zero. Indeed, consider the closed interval $[\omega_1', \omega_1'']$. Choose three points ω^1, ω^2, and ω^3 such that $\omega_1' < \omega^1 < \omega_1$, $\omega_1 < \omega^2 < \omega_2$, and $\omega_2 < \omega^3 < \omega_1''$. (We assume of course that $\omega_1' < \omega_1$ and $\omega_2 < \omega_1''$. The game G^n has no meaning otherwise.) Let F' be the strategy for nature which plays ω^1, ω^2, and ω^3 with equal probability. Clearly $\rho(F', \phi) \geqq \alpha > 0$ for all ϕ when $\alpha = \min(L_2(\omega^1), L_1(\omega^2), L_2(\omega^3))$. Hence $\nu_n \geqq \alpha > 0$ for all n.

Let $T_n = \{\omega : \omega_c[\omega_n', \omega_n''], \rho(\omega, \phi^n) = \nu_n\}$. T_n contains points in both preference regions; i.e., T_n contains ω in the interval (ω_1, ω_2) and T_n also meets at least one of the intervals (c, ω_1), (ω_2, d). Suppose not. Suppose $T_n \subset (\omega_1, \omega_2)$. $\bar{F}^n$ must concentrate at points of T_n. Let $\phi_0(x) \equiv 0$ for all x. Then $\rho(\bar{F}^n, \phi_0) = 0$ which contradicts the fact that $\nu_n > 0$. An analogous argument using $\phi_1(x) \equiv 1$ eliminates the possibility that T_n contain no points of (ω_1, ω_2).

The monotone minimax strategies ϕ^n are characterized by two points x_n, y_n $(x_n < y_n)$. In order to show that two subsequences of $\{\phi^n\}$ and $\{\bar{F}^n\}$ can be selected which converge to minimax strategies in the original game G we need the fact that the sets T_n are bounded away from the open ends or end of Ω. This is established by showing that the x_n's or y_n's or both are bounded away from c' and d', the ends of the spectrum of μ. c' may be $-\infty$ and d' may be $+\infty$.

Suppose Ω is open at c; Ω can be open or closed at d. We assert that c' cannot be a limit of the sequence $\{x_n\}$. Suppose there were a subsequence $\{x_{n_i}\}$ of $\{x_n\}$ with the limit c'.

Case 1. c' is a limit point of the sequence $\{y_{n_i}\}$.

There exists a subsequence $\{\phi^{n_j}\}$ of strategies such that $\{x_{n_j}\}$ and $\{y_{n_j}\}$ each have the limit c'. For $\omega < \omega_1$ and

$$\omega > \omega_2 \quad \rho(\phi^{n_j}, \omega) = L_2(\omega) \int_{x_{n_j}}^{y_{n_j}} P(x, \omega)\, d\mu(x), \qquad \text{Therefore } \rho(\phi^{n_j}, \omega) \to 0$$

as $n_j \to \infty$ for $\omega < \omega_1$ and $\omega > \omega_2$. Since $\nu_n \geqq \alpha > 0$ for all n, this means that T_n is totally contained in the interval (ω_1, ω_2), a contradiction.

Case 2. c' is not a limit point of $\{y_{n_i}\}$.

It is assumed there exists a $y_{c'}$ such that $y_{n_i} \geqq y_{c'}$ for all n_i.

LEMMA A. *Under the conditions that* $x_{n_i} \to c'$, *there exists a limit point of* $\{\nu_{n_i}\} \geqq l(c)$.

PROOF. By condition B given $\eta > 0$ there exists $M(\eta)$ such that $\int_{-\infty}^{y_c} P(x, \omega)\, d\mu(x) \geqq (1 - \eta)$ for $\omega < M(\eta)$. Given $\epsilon > 0$ there exists $\omega_\epsilon < M(\eta)$ such that $L_2(\omega_\epsilon) > l(c) - \epsilon$ (or if $l(c) = \infty$ for arbitrarily large K there exists $\omega_K < M(\eta)$ such that $L_2(\omega_K) > K$). For ω_ϵ (or ω_K) there exists $N(\omega_\epsilon, \epsilon')$ such that

$$\int_{-\infty}^{y_c} P(x, \omega_\epsilon)\, d\mu(x) - \int_{x_{n_i}}^{y_c} P(x, \omega_\epsilon)\, d\mu(x) < \epsilon'$$

for $n_i \geqq N(\omega_\epsilon, \epsilon')$. Therefore, when $l(c) < \infty$

$$\rho(\phi^{n_i}, \omega_\epsilon) \geqq L_2(\omega_\epsilon) \int_{x_{n_i}}^{y_c} P(x, \omega_\epsilon)\, d\mu(x) \geqq (l(c) - \epsilon)(1 - \eta - \epsilon')$$

for $n_i \geqq N(\omega_\epsilon, \epsilon')$. But ϵ, η, ϵ' are arbitrary constants so the assertion follows. When $l(c) = \infty$ $\rho(\phi^{n_i}, \omega_K) \geqq K(1 - \eta - \epsilon')$, and the assertion still holds.

LEMMA B. *Under the same conditions as in Lemma* A, $\nu_{n_i} \leqq l(c) - \beta$ *for all* n_i *where* $\beta > 0$.

PROOF. Let ϕ' be a monotone strategy for which the characterizing points x' and y' are interior to the spectrum of μ.

$$\rho(\omega, \phi') = L_1(\omega) \left[\int_{-\infty}^{x'} P(x, \omega)\, d\mu(x) + \int_{y'}^{\infty} P(x, \omega)\, d\mu(x) \right] + L_2(\omega) \int_{x'}^{y'} P(x, \omega)\, d\mu(x).$$

Let $a = l(c) - \max L_1(\omega)$. $a > 0$ and

$$L_1(\omega) \left[\int_{-\infty}^{x'} P(x, \omega)\, d\mu(x) + \int_{y'}^{\infty} P(x, \omega)\, d\mu(x) \right] \leqq l(c) - a.$$

If Ω is also open at d, then by virtue of conditions A and B for every $\epsilon > 0$ there exist two constants $H(\epsilon)$ and $K(\epsilon)$ such that $L_2(\omega) \int_{x'}^{y'} P(x, \omega)\, d\mu(x) < \epsilon$ for $\omega < H(\epsilon)$ and $\omega > K(\epsilon)$. For $H(\epsilon) \leqq \omega \leqq K(\epsilon)$ the second factor in $\rho(\omega, \phi')$ is $\leqq l(c) - b$ for some $b > 0$ by condition C. Thus $\rho(\omega, \phi') \leqq \max(l - a, \epsilon, l - b) < l$ for all ω, and $\nu_{n_i} \leqq \max(l - a, \epsilon, l - b)$ for all n_i. If Ω is closed at d, by condition B there exists $H(\epsilon)$ such that $L_2(\omega) \int_{x'}^{y'} P(x, \omega)\, d\mu(x) < \epsilon$ for $\omega < H(\epsilon)$. For $H(\epsilon) \leqq \omega \leqq d$ this factor is $\leqq l - b$ for some $b > 0$ by condition C. Again $\nu_{n_i} \leqq l - \beta$ where $\beta = l - \max(l - a, \epsilon, l - b) > 0$.

But Lemmas A and B are contradictory assertions. Hence the original assumption that c' was a limit point of $\{x_n\}$ is untenable.

An analogous argument shows that if Ω is open at d the sequence $\{y_n\}$ cannot have d' as a limit point.

The required theorem follows almost immediately.

LEMMA C. *If* Ω *is open at* c *and* d, *there exist two constants* C_1 *and* C_2 *such that*

$c < C_1 < C_2 < d$ and $T_n \subseteq [C_1, C_2]$ for all n. If Ω is open at $c(d)$ and closed at $d(c)$, there exists a constant $C_1(C_2)$ such that $c < C_1(C_2 < d)$ and $T_n \subseteq [C_1, d]$ ($[c, C_2]$) for all n.

PROOF. For ω near c and d $\rho(\omega, \phi^n) = L_2(\omega) \int_{x_n}^{y_n} P(x, \omega)\, d\mu(x)$. By the previous discussion if Ω is open at both ends there exist constants $x_{c'}$ and $y_{d'}$ interior to the spectrum of μ such that $x_{c'} \leqq x_n$ and $y_n \leqq x_{d'}$ for all n. $\rho(\omega, \phi^n) \leqq L_2(\omega) \int_{x_{c'}}^{y_{d'}} P(x, \omega)\, d\mu(x)$. By conditions A and B there exist constants $C_1(\eta)$ and $C_2(\eta)$ such that

$$L_2(\omega) \int_{x_{c'}}^{y_{d'}} P(x, \omega)\, d\mu(x) < \eta$$

for $c < \omega < C_1(\eta)$ and $C_2(\eta) < \omega < d$. Choose $\eta = \alpha/2$ where α is the bound of $\nu_n (\nu_n \geqq \alpha > 0$ for all $n)$. Thus T_n cannot have points below $C_1 = C_1(\alpha/2)$ or above $C_2 = C_2(\alpha/2)$. A similar argument works when Ω is open at just one end.

For each game $G^n = (\{F^n\}, \{\phi\}, \rho)$ in which Ω is open at both ends there is a triplet $(\bar{F}^n, \phi^n, \nu_n)$ where $\bar{F}^n$ is the minimax strategy for nature which concentrates at three points, ϕ^n is the monotone minimax strategy for the statistician characterized by two points, and ν_n is the value of the game. $\bar{F}^n$ concentrates at exactly three points (and not at at most three) for it has been shown that ϕ^n defines a split selection region for action 1, and by virtue of Theorem 5 and the fact that ϕ^n is Bayes against $\bar{F}^n$ this would be impossible unless $\bar{F}^n$ concentrates at three points. A subsequence $\{\phi^{n_i}\}$ can be selected which converges to a monotone strategy ϕ^* for the statistician. ϕ^* also defines a split selection region for action 1 since the x_{n_i}'s and y_{n_i}'s are bounded away from the ends of the spectrum of μ. Since the T_n's are contained in a closed interval contained in Ω, a subsubsequence $\{\bar{F}^{n_j}\}$ can be selected which converges to at most a three-point distribution, F^*, for nature. Finally a subsubsequence $\{\nu_{n_k}\}$ can be chosen which converges to a value ν. $\rho(F, \phi^{n_k}) \leqq \nu_{n_k}$ so by the Lebesgue convergence theorem $\rho(F, \phi^*) \leqq \nu$, for every F. Similarly $\rho(F^{n_k}, \phi) \geqq \nu_{n_k}$ so $\rho(F^*, \phi) \geqq \nu$. Thus ν is the value of the game, and F^* and ϕ^* are minimax strategies for nature and the statistician respectively. F^* concentrates at exactly three points since ϕ^* defines a split selection region for action 1.

When Ω is closed at one end, analogous arguments prove the existence of minimax strategies ϕ^* and F^* where F^* concentrates at at most three points.

Summing up the foregoing results, we have

THEOREM 8. *If conditions* A, B, *and* C *are satisfied and no other restriction on the parameter space Ω is made, then the game with payoff kernel $\rho(F, \phi)$ has a value, where optimal strategies exist with the same properties as is given in Theorem* 7.

Sec. 3. Computation of minimax strategies. The previous discussion has been an existence discussion, and no mention was made of how the statistician's monotone minimax strategy can be found or constructed. The remainder of this section is devoted to giving two general methods for constructing the minimax strategy—one for the one-sided problem and one for the two-sided problem.

In the one-sided problem

$$\rho(\omega, \phi) = \begin{cases} L_1(\omega) \int_{-\infty}^{x_0} P(x, \omega)\, d\mu(x), & \omega \geqq \omega_0 \\ L_2(\omega) \int_{x_0}^{\infty} P(x, \omega)\, d\mu(x), & \omega \leqq \omega_0 \end{cases}$$

where x_0 is the point characterizing the monotone strategy ϕ. As x_0 decreases $L_1(\omega) \int_{-\infty}^{x_0} P(x, \omega)\, d\mu(x)$ decreases for each $\omega \geqq \omega_0$ and $L_2(\omega) \int_{x_0}^{\infty} P(x, \omega)\, d\mu(x)$ increases for each $\omega \leqq \omega_0$. The method is now obvious. Choose an arbitrary x_0. If

$$\max_{\omega \leq \omega_0} \int_{x_0}^{\infty} P(x, \omega)\, d\mu(x) < \max_{\omega \geq \omega_0} \int_{-\infty}^{x_0} P(x, \omega)\, d\mu(x),$$

decrease x_0 until the maximums are equal. If the reverse is true, increase x_0. That x_0 which implies equal maximums above and below ω_0 defines the monotone minimax strategy. There is no danger of the maximums not existing since by conditions 1 and 2, $\rho(\omega, \phi) \to 0$ as $\omega \to c, d$.

In the two-sided problem

$$\rho(\omega, \phi) = \begin{cases} L_1(\omega) \left[\int_{-\infty}^{x_0} P(x, \omega)\, d\mu(x) + \int_{y_0}^{\infty} P(x, \omega)\, d\mu(x) \right] & \omega_1 \leqq \omega \leqq \omega_2 \\ L_2(\omega) \int_{x_0}^{y_0} P(x, \omega)\, d\mu(x) & \omega \leqq \omega_1, \omega \geqq \omega_2 \end{cases}$$

where x_0, y_0 are the points characterizing the monotone strategy ϕ. Assume Ω is open at both ends. Choose an arbitrary x_0. Determine y_0 (as a function of x_0) so that

$$\max_{\omega \leq \omega_1} L_2(\omega) \int_{x_0}^{y_0} P(x, \omega)\, d\mu(x) = \max_{\omega_2 \leq \omega} L_2(\omega) \int_{x_0}^{y_0} P(x, \omega)\, d\mu(x).$$

This cannot be done for all x_0. As $\omega \to d$, $L_2(\omega) \int_{x_0}^{d'} P(x, \omega)\, d\mu(x) \to l(d)$ and $l(d) > \max_{a \leq \omega \leq b} L_2(\omega)$ for any closed interval $[a, b] \subset \Omega$ so that for y_0 sufficiently large $\max_{\omega_2 \leq \omega} > \max_{\omega \leq \omega_1}$. Both maximums equal 0 for $y_0 = x_0$ so unless $\max_{\omega_2 \leq \omega} \geqq \max_{\omega \leq \omega_1}$, for all $y_0 > x_0$ there will be equality at some point. $\max_{\omega_2 \leq \omega} > \max_{\omega \leq \omega_1}$ for all $y_0 > x_0$ when x_0 is chosen too close to d'. In this case decrease x_0 until it is possible to determine y_0 so that the maximums are equal. There will be some point x_m such that for all $x_0 \leqq x_m$ a y_0 can be found. Now vary x_0 in the appropriate direction until the outer maximums are equal to

$$\max_{\omega_1 \leq \omega \leq \omega_2} \left[\int_{-\infty}^{x_0} P(x, \omega)\, d\mu(x) + \int_{y_0(x_0)}^{\infty} P(x, \omega)\, d\mu(x) \right].$$

The x_0 and corresponding y_0 which give three equal maximums determine the monotone minimax strategy for the statistician.

A further useful fact is that the monotone minimax strategy for the statistician is unique. This can be demonstrated with the aid of Lemma 4.

REFERENCES

[1] S. KARLIN, "Decision theory for Pólya type distributions; case of two actions, I." *Third Berkeley Symposium on Probability and Statistics*, Vol, 1, pp. 115–129.

[2] S. KARLIN AND H. RUBIN, "The theory of decision procedures for distributions with monotone likelihood ratio." *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272–299.

[3] Z. W. BIRNBAUM, "On random variables with comparable peakedness." *Ann. Math. Stat.*, Vol. 19 (1948), pp. 76–81.

[4] E. L. LEHMAN, *Notes on Testing Hypothesis*. University of California, 1950.

[5] H. CHERNOFF, "A property of some type A regions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 472–474.

[6] S. KARLIN, "The theory of infinite games," *Ann. Mathematics*, Vol. 24 (1953), pp. 371–401.

ON MINIMIZING AND MAXIMIZING A CERTAIN INTEGRAL WITH STATISTICAL APPLICATIONS[1, 2]

By Jagdish Sharan Rustagi

Carnegie Institute of Technology

1. Summary. We consider here the problem of minimizing and maximizing $\int_{-X}^{X}\varphi(x, F(x))\,dx$ under the assumptions that $F(x)$ is a cumulative distribution function (cdf) on $[-X, X]$ with the first two moments given and that φ is a certain known function having certain properties. The existence of the solution has been proved and a characterization of the maximizing and minimizing cdf's given. The minimizing cdf is unique when $\varphi(x, y)$ is strictly convex in y and is completely characterized for some special forms of φ. The maximizing cdf is a discrete distribution and in the above case turns out to be a three-point distribution. Several statistical applications are discussed.

2. Introduction. Let $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ be n ordered independent observations from a population with cdf $F(x)$ having standard deviation σ. Let $w_n = x_n - x_1$ denote the sample range. Then it is well-known that

$$E(w_n) = \int_{-\infty}^{\infty} x\, d\{F^n(x) + (1 - F(x))^n\} \tag{2.1}$$

and

$$E(x_n) = \int_{-\infty}^{\infty} x\, d\{F^n(x)\}. \tag{2.2}$$

Plackett [9] considered the problem of establishing universal upper and lower bounds for $[E(w_n)]/\sigma$ on the lines of Chebycheff inequalities for moments. Moriguti [14] considered an equivalent case of establishing bounds for $E(x_n)$, but he assumed that the underlying distribution is symmetrical.

Gumbel [10] uses a variational method to derive the solution of the problem of maximizing $E(w_n)$ and $E(x_n)$ over the class of continuous cdf's with given mean and variance and gives a sort of sufficiency condition. Hartley and David [1] consider the same problem of maximizing $E(x_n)$ as in [10], and obtain the solution of the problem of maximizing and minimizing $E(w_n)$ but they assume, in addition, that $F(x)$ is a cdf on the bounded range $[-X, X]$.

Integrating (2.1) and (2.2) by parts, we find that the problems of maximizing (minimizing) $E(w_n)$ and $E(x_n)$ are the same as those of minimizing (maximizing)

Received April 6, 1956; revised July 17, 1956.

[1] This paper is based on the author's doctoral dissertation, accepted by Stanford University.

[2] This research was done under the sponsorship of the Office of Naval Research.

$$\int_{-X}^{X} [F(x)^n + (1 - F(x))^n]\, dx \qquad \text{and} \qquad \int_{-X}^{X} F(x)^n\, dx,$$

respectively, with appropriate restrictions on $F(x)$. We see here that $\varphi(x, y) = y^n + (1 - y)^n$ or $\varphi(x, y) = y^n$ is strictly convex in y for $0 \leqq y \leqq 1$.

There are many other situations in statistics where problems of maximizing and minimizing an integral of a strictly convex function of $F(x)$ occur. In the evaluation of efficiencies of various nonparametric tests of hypotheses, we are faced with integrals of the above type. For example, Birnbaum and Klose [7] have derived a lower bound for the variance of the Mann-Whitney Statistic—an improvement on the lower bound due to V. Dantzig, which is based on minimizing $\int_0^1 [F(x) - x]^2\, dx$, where $F(x)$ is a cdf on $0 \leqq x \leqq 1$ and $\int_0^1 F(x)\, dx = 1 - p \geqq \frac{1}{2}$. Here again $[F(x) - x]^2$ is strictly convex in $F(x)$.

The above problems suggest a generalization.

We consider the problem of maximizing and minimizing an integral of $\varphi(x, F(x))$, where $\varphi(x, y)$ is strictly convex in y. A very special case of this is the one where $\varphi(x, y)$ is a function of y alone and includes the important applications of minimizing and maximizing $E(w_n)$, $E(x_n)$, etc. Many other authors, to name a few such as Chernoff and Reiter [2], Rubin and Isaacson [13], Karlin [3, 4], Hoeffding [11], Hoeffding and Shrikhande [15], Brunk, Ewing and Utz [6], have also considered related problems.

We have used Karlin's [3, 4] technique in the solution of the minimum problem. We compare our technique with that of [2] and [3] in Section 6 and obtain results similar to those in [2] and [3]. The maximum problem is discussed in Section 8. We find that the maximizing cdf is a discrete distribution and in our case a three-point distribution. David and Hartley [1] have shown further that the minimizing cdf for $E(w_n)$ with given restrictions on mean and variance is a two-point distribution. This does not seem to be true in general. Similar results were obtained in [2], [11], and [15].

The results and techniques of our paper have many other applications besides those discussed above. Many classical inequalities of the Chebycheff type can be obtained with the help of our results. In Section 7 we discuss an example where the techniques of this paper yield the solution to a problem of a different type.

3. Statement of the problem and existence of its solution. Let $S = \{(x, y): -X \leqq x \leqq X, 0 \leqq y \leqq 1\}$ where X is already specified.

Let φ be a function defined on the closed and bounded region S such that

(1) φ is bounded and continuous in S,

(2) φ is strictly convex and twice differentiable in y.

We shall minimize (maximize)

$$I(F) = \int_{-X}^{X} \varphi(x, F(x))\, dx \tag{3.1}$$

over all $F \varepsilon \mathfrak{A}$ where $\mathfrak{A}$ is the class of all admissible cdf's, i.e., cdf's satisfying the following constraints:

$$\int_{-X}^{X} x\, dF(x) = \mu_1, \qquad \int_{-X}^{X} x^2\, dF(x) = \mu_2, \tag{3.2}$$

and

$$F(x) = \begin{cases} 0, & x < -X, \\ 1, & x > X. \end{cases}$$

Here μ_1 and μ_2 are such that $\mu_2 > \mu_1^2$ and $\mu_2 < X^2$. In this case there exist cdf's satisfying (3.2) and hence class $\mathfrak{A}$ is non-null.

Integrating by parts the integrals in (3.2), the restrictions become

$$\begin{aligned} \int_{-X}^{X} F(x)\, dx &= X - \mu_1, \\ \int_{-X}^{X} xF(x)\, dx &= \frac{X^2 - \mu_2}{2}. \end{aligned} \tag{3.3}$$

We shall first show that an admissible minimizing (maximizing) cdf exists. Let $\mathfrak{F}$ be the class of all cdf's defined on $[-X, X]$. Then the following is well-known [8]:

LEMMA 3.1. *$\mathfrak{F}$ is convex and compact in the topology of convergence in distribution.* (The compactness of Lemma 3.1 is a restatement of the Helly-Bray lemma.)

Define a transformation

$$T: \mathfrak{F} \to R \qquad \text{such that}$$

$$T \circ F = \left(\int_{-X}^{X} \varphi(x, F(x))\, dx, \int_{-X}^{X} F(x)\, dx, \int_{-X}^{X} xF(x)\, dx \right).$$

It is easy to see that T is a continuous transformation as $\varphi(x, y)$ is continuous in y. But a continuous transformation maps a compact set into a closed and bounded set [8]. Hence we have the following lemma.

LEMMA 3.2. *The set Γ of points*

$$\left(\int_{-X}^{X} \varphi(x, F(x))\, dx, \int_{-X}^{X} F(x)\, dx, \int_{-X}^{X} xF(x)\, dx \right), \quad \text{for} \quad F \varepsilon \mathfrak{F},$$

is a closed and bounded set in R.

The restrictions (3.3) define a cross section Γ_1 of a closed and bounded set Γ, and hence Γ_1 is also closed and bounded. Therefore, the minimizing and maximizing points exist and are given by the boundary points of Γ_1 so long as Γ_1 is non-null. But Γ_1 is non-null as $\mathfrak{A}$ is non-null as seen before. Hence the minimizing and maximizing admissible cdf's exist.

4. Reduction of the minimum problem to subsidiary problems and uniqueness of its solution. In this section we first prove the uniqueness of the minimizing

cdf, using the property of strict convexity of the function φ in its second argument. In characterizing the solution of the minimizing problem, we use Karlin's method [3] to reduce the main problem of minimizing the integral (3.1) over the class $\mathfrak{A}$ of admissible cdf's, to a subsidiary problem of minimizing an integral of a related function over all cdf's $\mathfrak{F}$. This reduction together with the uniqueness of the minimizing cdf gives us a characterization of the minimizing cdf which we give in the next section.

LEMMA 4.1. *There is a unique cdf F_0, which minimizes* (3.1) *subject to the side conditions* (3.3) *when $\varphi(x, y)$ is strictly convex in y.*

PROOF. Suppose the solution is not unique. Let $F_0(x)$ and $F_1(x)$ be two distinct admissible cdf's which minimize (3.1). Let

$$M = \min_{F \in \mathfrak{A}} \int_{-X}^{X} \varphi(x, F(x))\, dx.$$

As φ is strictly convex in y, for $0 < \lambda < 1$,

$$\begin{aligned} &\int_{-X}^{X} \varphi(x, \lambda F_0(x) + (1 - \lambda)F_1(x))\, dx \\ &\qquad < \lambda \int_{-X}^{X} \varphi(x, F_0(x)\, dx + (1 - \lambda) \int_{-X}^{X} \varphi(x, F_1(x)\, dx \\ &\qquad = \lambda M + (1 - \lambda)M \\ &\qquad = M. \end{aligned} \tag{4.0}$$

But M is the minimum, and hence we have a contradiction.

We shall now prove the following lemmas, with the help of which we shall reduce the main problem to a simpler problem.

LEMMN 4.2. *$F_0(x)$ minimizes* (3.1) *if and only if*

$$\int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, y)\,|_{y=F_0(x)} F(x)\, dx \geqq \int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, y)\,|_{y=F_0(x)} F_0(x)\, dx \tag{4.1}$$

for all $F \varepsilon \mathfrak{A}$.

PROOF. For any other admissible cdf $F(x)$, define

$$I(\lambda) = \int_{-X}^{X} \varphi(x, \lambda F_0(x) + (1 - \lambda)F(x))\, dx, \qquad 0 \leqq \lambda \leqq 1.$$

As φ is twice differentiable in y, $\partial\varphi/\partial y$ exists and is continuous in y, and hence $I(\lambda)$ is differentiable and is given by

$$I'(\lambda) = \int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, \lambda F_0(x) + (1 - \lambda)F(x))(F_0(x) - F(x))\, dx.$$

Since φ is strictly convex in y, it follows very easily that $I(\lambda)$ is a strictly convex function of λ. If $F_0(x)$ minimizes (3.1), then $I(\lambda)$ achieves its minimum at $\lambda = 1$; and this is possible if and only if

$$I'(\lambda)\,|_{\lambda=1} \leqq 0,$$

i.e.,

$$\int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, y)\,|_{y=F_0(x)}[F_0(x) - F(x)]\,dx \leqq 0,$$

or,

$$\int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, y)\,|_{y=F_0(x)}F_0(x)\,dx \leqq \int_{-X}^{X} \frac{\partial}{\partial y}(x, y)\,|_{y=F_0(x)}F(x)\,dx.$$

Conversely let (4.1) hold true. Then we have $I'(\lambda)\,|_{\lambda=1} \leqq 0$, and hence by the strict convexity of $I(\lambda)$ we have $I(1) < I(0)$ or,

$$\int_{-X}^{X} \varphi(x, F_0(x))\,dx < \int_{-X}^{X} \varphi(x, F(x))\,dx,$$

i.e., $F_0(x)$ minimizes (3.1). This proves the lemma.

We use the following notation:

$$I_{F_0}(F) = \int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, F_0(x))F(x)\,dx.$$

With the help of the above lemma, we find that the problem P_1 of minimizing $I(F)$ over all admissible cdf's, is related to the problem P_{2F_0} of finding an admissible $F(x)$ which minimizes $I_{F_0}(F)$. In fact, we are interested in finding an F_0 such that F_0 is a solution of P_{2F_0}. This looks like a complicated problem, but we now have a problem linear in F which is relatively easy to deal with.

Because P_1 has a unique solution, Lemma 4.2 implies that there is one and only one F_0 such that F_0 solves P_{2F_0}. This, however, does not mean that P_{2F_0} has a unique solution.

Let $T:\mathfrak{F} \to \Gamma_2$ be a transformation given by

$$T \circ F = \left(\int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, F_0(x))F(x)\,dx, \int_{-X}^{X} F(x)\,dx, \int_{-X}^{X} xF(x)\,dx\right).$$

Obviously T is bounded and is linear and hence continuous in F. But as $\mathfrak{F}$ is convex and compact in the topology of convergence in distribution by Lemma 3.1, the transformation T maps the convex and compact set into a convex and compact set Γ_2, and hence we have the following result.

LEMMA 4.3. *Γ_2 is a convex, closed and bounded set in three dimensions.*

Solving P_{2F_0} corresponds to finding the minimum among all points of Γ_2 for which

$$\int_{-X}^{X} F(x)\,dx = X - \mu_1,$$

$$\int_{-X}^{X} xF(x)\,dx = \frac{X^2 - \mu_2}{2},$$

and this will be a boundary point of the set Γ_2.

Suppose F_0 solves P_{2F_0}. Then F_0 corresponds to a boundary point of Γ_2, and there is a supporting hyperplane of Γ_2 at the minimum point $\gamma = (u_0, v_0, w_0)$, i.e., for some η_0, η_1, η_2 and $\eta_3(\eta_0, \eta_1, \eta_2$ not all zero),

$$\eta_0 u_0 + \eta_1 v_0 + \eta_2 w_0 + \eta_3 = 0, \tag{4.2}$$

and

$$\eta_0 u + \eta_1 v + \eta_2 w + \eta_3 \geqq 0 \tag{4.3}$$

for all other points (u, v, w) belonging to Γ_2, where

$$u = \int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, F_0(x)) F(x)\, dx,$$

$$v = \int_{-X}^{X} F(x)\, dx, \qquad w = \int_{-X}^{X} xF(x)\, dx,$$

therefore,

$$\eta_0(u - u_0) + \eta_1(v - v_0) + \eta_2(w - w_0) \geqq 0. \tag{4.5}$$

We shall see below that η_0 can be taken positive and hence can be normalized so as to be equal to one. Therefore, by taking $\eta_0 = 1$ in (4.5) we have

$$\int_{-X}^{X} \left[\frac{\partial \varphi}{\partial y}(x, F_0(x)) + \eta_1 + \eta_2 x\right] F(x)\, dx \geqq \int_{-X}^{X} \left[\frac{\partial \varphi}{\partial y}(x, F_0(x)) + \eta_1 + \eta_2 x\right] F_0(x)\, dx.$$

Hence $F_0(x)$ minimizes

$$\int_{-X}^{X} \left[\frac{\partial \varphi}{\partial y}(x, F_0(x)) + \eta_1 + \eta_2 x\right] F(x)\, dx \tag{4.6}$$

among the class $\mathfrak{F}$ of all cdf's on $[-X, X]$.

Conversely, if $F_0(x) \in \mathfrak{A}$ minimizes (4.6), we have, retracing the steps, that $(u - u_0) + \eta_1(v - v_0) + \eta_2(w - w_0) \geqq 0$. Suppose $F(x)$ is admissible, and hence $v = v_0$ and $w = w_0$, and hence $u - u_0 \geqq 0$, i.e.,

$$\int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, F_0(x)) F(x)\, dx \geqq \int_{-X}^{X} \frac{\partial \varphi}{\partial y}(x, F_0(x)) F_0(x)\, dx.$$

In other words, F_0 minimizes $I_{F_0}(F)$ over all admissible cdf's $\mathfrak{A}$.

We shall now show that η_0 can be taken positive. Let

$$\Gamma_2^* = \{(u^*, v, w): u^* \geqq u, (u, v, w) \in \Gamma_2\}.$$

Then Γ_2^* is obviously convex and $\Gamma_2 \subseteq \Gamma_2^*$. u_0 is the minimum of u subject to the conditions that $v = v_0$ and $w = w_0$. This implies that (u_0, v_0, w_0) is also a minimum point of Γ_2^* and hence is its boundary point. Hence there is an $(\eta_0, \eta_1, \eta_2) \neq (0, 0, 0)$ such that

$$\eta_0(u^* - u_0) + \eta_1(v - v_0) + \eta_2(w - w_0) \geqq 0 \tag{4.7}$$

for points (u^*, v, w) belonging to Γ_2^*. Hence

$$\eta_0(u - u_0) + \eta_1(v - v_0) + \eta_2(w - w_0) \geqq 0$$

for $(u, v, w) \varepsilon \Gamma_2$. Suppose $\eta_0 = 0$. Then we have

$$\eta_1(v - v_0) + \eta_2(w - w_0) \geqq 0,$$

or

$$\int_{-X}^{X} (\eta_1 + \eta_2 x)F(x)\,dx \geqq \int_{-X}^{X} (\eta_1 + \eta_2 x)F_0(x)\,dx,$$

i.e., F_0 minimizes $\int_{-X}^{X} (\eta_1 + \eta_2 x)F(x)\,dx$ over all $F \varepsilon \mathfrak{F}$. Now $\eta_1 + \eta_2 x$ is either nondecreasing or nonincreasing according as $\eta_2 \geqq 0$ or $\eta_2 \leqq 0$, and hence the unique minimizing cdf of the above integral is a two-point distribution with its total mass concentrated at $-X$ and X so that $\mu_2 = X^2$. But such a cdf is not admissible and hence there is a contradiction.

It is easily seen now that η_0 is not negative. Suppose η_0 is negative. Consider then a point $(u_0 + h, v_0, w_0) \varepsilon \Gamma_2^*$ for some $h > 0$, so that from (4.7) we obtain

$$\eta_0 h \geqq 0,$$

which is again a contradiction. Hence η_0 is positive.

REMARK. Another way to show that $\eta_0 \neq 0$ would be as follows: $\eta_0 = 0$ corresponds to boundary points of the set Γ_2 where the supporting hyperplanes are parallel to the u-axis, and hence (v_0, w_0) corresponds to the boundary of the projection Γ_2 on the (v, w) plane. But the conditions on the first two moments are such that the given point (v_0, w_0) will be interior to the projection set, and hence $\eta_0 \neq 0$.

The previous argument applies only in the special case of the first two moments of $F(x)$ being given. In general when more moments are specified, the latter argument will apply if we impose conditions on the given moments such that the given point is interior to the moment space which is analogous to the projection of the set Γ_2. It easily follows then that $\eta_0 > 0$, in general. Let

$$I_{F_0\eta_1\eta_2}(F) = \int_{-X}^{X} \left[\frac{\partial \varphi}{\partial y}(x, F_0(x)) + \eta_1 + \eta_2 x\right] F(x)\,dx,$$

and let the problem $P_{2F_0\eta_1\eta_2}$ be that of finding the minimum of $I_{F_0\eta_1\eta_2}(F)$ over all $F(x) \varepsilon \mathfrak{F}$.

The above results are summarized in the following lemma.

LEMMA 4.4. *F_0 solves P_{2F_0} if F_0 is an admissible cdf which solves $P_{3F_0\eta_1\eta_2}$, and any F_0 which solves P_{2F_0} solves $P_{3F_0\eta_1\eta_2}$ for some η_1 and η_2.*

5. Characterization of the solution. In this section we characterize the solution of the minimum problem in terms of $f_{\eta_1\eta_2}(x)$ which is that value of y for which

$$\frac{\partial \varphi}{\partial y}(x, y) + \eta_1 + \eta_2 x = 0.$$

Let

$$A(x) = B_{\eta_1\eta_2}(x, F_0(x)) = \frac{\partial \varphi}{\partial y}(x, F_0(x)) + \eta_1 + \eta_2 x.$$

Since $\frac{\partial \varphi}{\partial y}(x, y)$ is continuous in y, $A(x)$ can have a discontinuity only if $F_0(x)$ has a jump. But as $\frac{\partial \varphi}{\partial y}(x, y)$ is increasing in y, the discontinuities of $A(x)$ are upward jumps. Also since φ is continuous in the region

$$S = \{(x, y): -X \leqq x \leqq X, 0 \leqq y \leqq 1\},$$

$A(x)$ is bounded in S. We then have the following theorems which will characterize the solution of our problems.

THEOREM 5.1. *If F_0 solves $P_{\eta_1\eta_2}$, the set $\{x: A(x) \neq 0, -X < x < X\}$ has F_0-measure zero.*

PROOF. Suppose that F_0 is continuous on the right. Consider the set $S_p = \{x: A(x) > 0, -X < x < X\}$. It is a denumerable union of intervals $[x_1, x_2)$. We shall show that

$$F_0(x_2) = F_0(x_1 -)$$

and therefore, the interval $[x_1, x_2)$ has F_0-measure zero. Suppose this were not the case. Then as $A(x) > 0$ and is increasing in y, so that

$$\int_{x_1}^{x_2} A(x)F_0(x_1 -)\, dx < \int_{x_1}^{x_2} A(x)F_0(x)\, dx,$$

there is a contradiction. Consequently, S_p has F_0-measure zero.

Consider now the set $S_q = \{x: A(x) < 0, -X < x < X\}$. Because all discontinuities are upward jumps and $A(x)$ is continuous on the right, S_q is an open set. Hence S_q is denumerable union of intervals (x_1, x_2). Then we shall see that $F_0(x_2) = F_0(x_1) = 0$ and that the F_0-measure of the interval (x_1, x_2) is zero. We also prove this by contradiction, as otherwise,

$$\int_{x_1}^{x_2} A(x)F_0(x_2)\, dx < \int_{x_1}^{x_2} A(x)F_0(x)\, dx.$$

Since S_q is a denumerable union of such intervals, S_q has also F_0-measure zero.

Hence the above arguments show that the set $\{x: A(x) \neq 0, -X < x < X\}$ has F_0-measure zero.

REMARKS. 1. It is easy to see that if $A(-X) > 0$, then $F_0(-X) = 0$ and if $A(X) < 0$, F_0 is continuous at X.

2. The following corollary shows that the integral of $A(x)$ is zero over intervals on which F is constant.

COROLLARY. *If $F_0(x)$ be such that $F_0(x) = c$, $0 < c < 1$ for $a \leqq x < b$ and $F_0(x) < c$ for $x < a$, $F_0(x) > c$ for $x > b$, then*

$$\int_a^b A(x)\, dx = 0.$$

PROOF. Suppose $\int_a^b A(x)\,dx < 0$ and $b < X$. Replace $F_0(x)$ on the interval $[a, b + \delta)$ for any small number $\delta > 0$, by the constant quantity $F_0(b + \delta)$. Let ν be the increase in $I_{F_0\eta_1\eta_2}(F)$ due to this replacement. Then

$$\begin{aligned}\nu &= \int_a^{b+\delta} A(x)F_0(b_i + \delta)\,dx - \int_a^{b+\delta} A(x)F_0(x)\,dx \\ &= (F_0(b + \delta) - c)\int_a^{b+\delta} A(x)\,dx - \int_a^{b+\delta} A(x)[F_0(x) - c]\,dx \\ &\leqq [F_0(b + \delta) - c]\left(\int_a^b A(x)\,dx + 2M\delta\right), \quad \text{where} \quad |A(x)| < M.\end{aligned}$$

Letting $\delta \to 0$, we find that ν becomes negative and hence there is a contradiction. The case where $b = X$ is trivial.

If we suppose that $\int_a^b A(x)\,dx > 0$ and a is a point of continuity of F_0, then by an argument similar to that above, we get the contradiction when $a > -X$, by replacing F_0 on $(a - \delta, b)$ by $F_0(a - \delta)$ and letting $\delta \to 0$. In case $a = -X$ or there is a jump in F_0 at a, the proof is trivial.

REMARK. If $K(x)$ is a function satisfying the properties of the function $A(x)$, then the problem P_K (corresponding to $P_{3F_0\eta_1\eta_2}$) of finding a cdf F_0 such that

$$I_K(F_0) = \min_{F\in\mathfrak{F}} I_K(F) = \min_{F\in\mathfrak{F}} \int_{-X}^{X} K(x)F(x)\,dx,$$

gives the same results as stated in Theorem 5.1 and its corollary, i.e., if F_0 is the solution of P_K,

(a) the set $\{x: K(x) \neq 0, -X < x < X\}$ is of F_0-measure zero.

(b) $\int K(x)F_0(x)\,dx$ is zero over intervals where F_0 is constant.

THEOREM 5.2. *If F_0 solves $P_{3F_0\eta_1\eta_2}$, then F_0 has no jumps on the open interval $(-X, X)$ and hence $A(x)$ is continuous on $(-X, X)$.*

PROOF. Let F_0 have a jump at x_0, $-X < x_0 < X$. Then by Theorem 5.1, $A(x_0) = 0$. But since $\partial\varphi/\partial y$ is strictly increasing in y, x_0 is the right-hand end-point of an interval on which $A(x) < 0$. By the same arguments as in the proof of the Theorem 5.1, we see that on this interval $F_0(x) = F_0(x_0)$, and hence F_0 has no discontinuity. But because discontinuities of $A(x)$ arise on account of jumps of F_0, there are no discontinuities in $A(x)$, or $A(x)$ is continuous on the interval $(-X, X)$.

Let $f_{\eta_1\eta_2}(x)$ be defined with $0 \leqq f_{\eta_1\eta_2}(x) \leqq 1$ such that $B_{\eta_1\eta_2}(x, f_{\eta_1\eta_2}(x)) = 0$. (The function $f_{\eta_1\eta_2}$ is defined on that subset of $[-X, X]$ for which there exists a y between 0 and 1 such that $B_{\eta_1\eta_2}(x, y) = 0$.)

As $\partial\varphi(x, y)/\partial y$ is continuous and strictly increasing in y, $f_{\eta_1\eta_2}(x)$ is continuous wherever it is defined. If $0 < f_{\eta_1\eta_2}(x_0) < 1$, then $f_{\eta_1\eta_2}$ is defined in some interval about x_0 (the interval is one-sided if $x_0 = \pm X$). Graphically $f_{\eta_1\eta_2}$ represents a number of curve segments which terminate when $f_{\eta_1\eta_2}(x)$ is zero or one.

More specifically, $f_{\eta_1\eta_2}(x)$ is defined on the union of closed intervals at the end-points of such of which it is either zero or one. Let $[a_i, b_i]$ and $[a_j, b_j]$ be

two such intervals, not separated by any others such that $b_i \leqq a_j$, then $f_{\eta_1\eta_2}(b_i) = f_{\eta_1\eta_2}(a_j)$. If there are an infinite number of intervals $[a_j, b_j]$ in the neighborhood of b_i, it follows that

$$f_{\eta_1\eta_2}(b_i) = f_{\eta_1\eta_2}(a_j) = f_{\eta_1\eta_2}(b_j)$$

for b_j sufficiently close to b_i. Hence the following definition of a function $g_{\eta_1\eta_2}$ has a meaning.

DEFINITION. Define $g_{\eta_1\eta_2}$ to be that unique function on $[-X, X)$ which is continuous on $[-X, X)$ such that

$$g_{\eta_1\eta_2}(x) = \begin{cases} f_{\eta_1\eta_2}(x), & \text{where } f_{\eta_1\eta_2} \text{ is defined,} \\ 0 \text{ or } 1, & \text{elsewhere,} \end{cases}$$

and

$$g_{\eta_1\eta_2}(X) = 1,$$

provided that the subset of $[-X, X)$ for which $f_{\eta_1\eta_2}$ is defined, is non-null.

THEOREM 5.3. *If F_0 solves $P_{3F_0\eta_1\eta_2}$, then for $-X \leqq x < X$, F_0 coincides with $g_{\eta_1\eta_2}$ except on intervals on which F_0 is constant.*

PROOF. From Theorems 5.1 and 5.2, we know that F_0 has no jumps on $(-X, X)$ and F_0 cannot increase when $A(x) \neq 0$. Therefore, F_0 remains constant until it intersects with $f_{\eta_1\eta_2}$.

COROLLARY. *If $g_{\eta_1\eta_2}$ is a cdf, then $F_0(x) = g_{\eta_1\eta_2}(x)$.*

REMARKS. 1. We can represent a conceivable situation by Fig. 1.

2. It must be noted that the corollary to Theorem 5.1 puts a strong restriction on the intervals on which F_0 is constant.

3. The solution in the general case may not be completely specified, but we shall consider in the following some special cases where the minimizing cdf is completely characterized.

Special Cases.

I. When $\partial\varphi(x, y)/\partial y$ is nonincreasing in x.

THEOREM 5.4. *If $\partial\varphi(x, y)/\partial y$ is nonincreasing in x and $\eta_2 < 0$, then $F_0(x) = g_{\eta_1\eta_2}(x)$ for $-X < x < X$.*

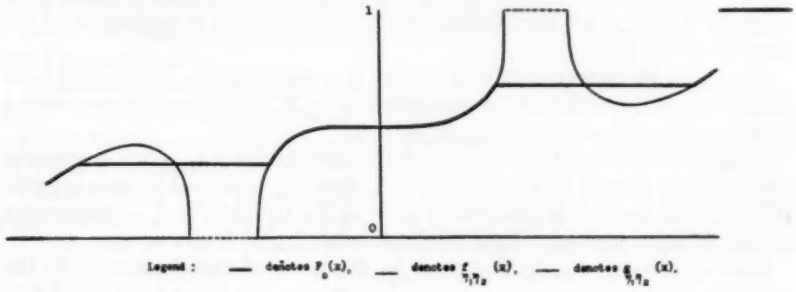


FIG. 1

PROOF. If the conditions of the theorem hold, $A(x)$ is a decreasing function of x and hence $f_{\eta_1\eta_2}$ is increasing in x so that $g_{\eta_1\eta_2}$ is a cdf. We get the result of the theorem, then, by the corollary to Theorem 5.3.

II. When φ is a function of y alone, i.e., $\varphi(x, y) = \psi(y)$.

LEMMA 5.1. *If $\varphi(x, y) = \psi(y)$, then corresponding to F_0, which is the solution of P_1, η_2 is negative.*

PROOF. If $\eta_2 \geqq 0$, $\eta_1 + \eta_2 x$ is nondecreasing. Also as $\psi'(y)$ is nondecreasing in x, the function $A(x) = \psi'(F_0) + \eta_1 + \eta_2 x$ is nondecreasing in x. Therefore, $f_{\eta_1\eta_2}(x)$ is nonincreasing, and hence from Theorem 5.3, it follows that F_0 is constant on $[-X, X)$. But for such F_0, $\mu_2 = X^2$, and hence F_0 is not admissible. Therefore, $\eta_2 < 0$.

Theorem 5.4 and Lemma 5.1 imply, then, the following theorem.

THEOREM 5.5. *If $\varphi(x, y) = \psi(y)$, the solution F_0 of P_1 is given by $g_{\eta_1\eta_2}$ for some η_1, η_2.*

REMARK. Unfortunately it is not always true that $\eta_2 < 0$ as assumed in Theorem 5.4. In fact for side conditions corresponding to small variance, one has $\eta_2 > 0$. It might still happen that $f_{\eta_1\eta_2}$ is nondecreasing, and then the result of Theorem 5.4 still holds. In any case Theorem 5.3 with the corollary to Theorem 5.1 gives a useful characterization of the solution of our problem.

6. Comparison of our Technique with that of Gumbel [10] and Chernoff and Reiter [2].

(i) *Gumbel's Method.* The problem considered by Gumbel is that of maximizing

$$\int_0^1 x(F) n F^{n-1}\, dF, \tag{6.1}$$

with restrictions

$$\int_0^1 x(F)\, dF = 0, \qquad \int_0^1 x^2(F)\, dF = 1.$$

A variational technique has been used to derive the form obtained for the maximizing cdf is given by equating to zero, the first variation of

$$\int \{n x(F) F^{n-1} + \eta_1 x(F) + \eta_2 x^2(F)\}\, dF, \tag{6.2}$$

i.e.,

$$n F^{n-1} + \eta_1 + 2\eta_2 x(F) = 0, \qquad 0 \leqq F \leqq 1. \tag{6.3}$$

The above equation gives a sort of sufficiency condition as any admissible F given by (6.3) does maximize the integral (6.2) and hence maximizes (6.1). David and Hartley [1] have given an ingenious argument to prove the sufficiency of the solution, but that seems unnecessary. However, the above equation does not give the necessity of the solution, since this approach does not provide an argument for proving that the constants η_1 and η_2, to make the cdf admissible, always exist.

This method also extends to the case of a bounded random variable as treated in this paper.

We shall use the above approach for our problem. Integrating by parts, we have

$$\int_{-x}^{x} \varphi(x, F(x))\, dx = x\varphi(x, F(x))\Big]_{-x}^{x} - \int_{-x}^{x} x \frac{d\varphi}{dx}(x, F)\, dx.$$

Now

$$d\varphi(x, F(x)) = \frac{\partial \varphi}{\partial x}(x, F(x))\, dx + \frac{\partial \varphi}{\partial y}(x, F(x))\, dF.$$

When $\varphi(x, F(x))$ is a function of F alone, say, $\psi(F(x))$, $d\psi = \psi'(F)\, dF$, and hence the problem of minimizing

$$\int_{-x}^{x} \psi(F)\, dx$$

is the same as that of maximizing

$$\int_{0}^{1} x(F)\psi'(F)\, dF.$$

Hence using Gumbel's approach, we maximize

$$\int_{0}^{1} [x(F)\psi'(F) + \eta_1 x(F) + \eta_2 x^2(F)]\, dF$$

and get the following equation satisfied by the admissible maximizing cdf,

$$\psi'(F) + \eta_1 + 2\eta_2 x(F) = 0.$$

In the above case, our technique would also lead to a similar equation. But in the general case when we consider $\varphi(x, F(x))$, Gumbel's approach does not seem to apply.

(ii) *Chernoff and Reiter Method.* Chernoff and Reiter [2] consider the problem of minimizing and maximizing

$$\int g(x)\, dF(x),$$

with side conditions

$$\int x\, dF(x) = c_1, \qquad \int x^2\, dF(x) = c_2$$

such that $c_2 > c_1^2$ and $g(x)$ is a continuous function of x.

In the process of reduction of our main problem, we have an intermediary problem P_{2F_0} of finding the minimum of

$$I_{F_0}(F) = \int_{-x}^{x} \frac{\partial \varphi}{\partial y}(x, F_0(x))F(x)\, dx$$

over all admissible cdf's $\mathfrak{A}$. Now as

$$f(x) = \int_{-X}^{x} \frac{\partial \varphi}{\partial y}(x, F_0(x))\, dx$$

is continuous in x, integrating by parts $I_{F_0}(F)$, we have

$$I_{F_0}(F) = \text{constant} - \int_{-X}^{X} f(x)\, dF(x),$$

or we maximize

$$\int_{-X}^{X} f(x)\, dF(x)$$

over all admissible cdf's $\mathfrak{A}$. Now as $f(x)$ is continuous, by the methods of Chernoff and Reiter, the necessary condition for the maximum is given by the following.

(a) There is an η_1 and η_2 such that when x is a point of continuity of $F(x)$,

$$B_{\eta_1\eta_2}(x, F_0(x)) = \frac{\partial \varphi}{\partial y}(x, F_0(x)) + \eta_1 + 2\eta_2 x = 0,$$

except on a set of F_0-measure zero

(b) F_0 has no jump in $-X < x < X$, otherwise either $B_{\eta_1\eta_2}(x, F_0(x)) > 0$ or $B_{\eta_1\eta_2}(x, F_0(x-)) < 0$. Hence we get a result similar to our result obtained in Section 5, i.e., the set

$$\{x : B_{\eta_1\eta_2}(x, F_0(x)) \neq 0, \quad -X < x < X\}$$

has F_0-measure zero.

7. Examples. In this section we discuss some examples to illustrate the method of obtaining the minimizing cdf for our problem for some specified function φ. We also discuss an example of the special case $\varphi(x, y) = \psi(y)$. We have included an example where the methods of the paper have been used to solve a problem of a different type.

EXAMPLE 1. Consider the problem of finding

$$\min_{F \in \mathfrak{A}} \frac{1}{2} \int_{-X}^{X} [F(x) - x]^2\, dx,$$

when $\mathfrak{A}$ denotes the admissible class of cdf's as in Section 3.

This is the special case of a more general problem where we minimize $\int_{-X}^{X} \psi(F(x) - x)\, dx$ for $F \in \mathfrak{A}$. Here $\psi(y - x) = \frac{1}{2}(y - x)^2$, ψ being a strictly convex, bounded, and continuous function of its argument. This problem has also been discussed by Birnbaum and Klose [7], as a lemma to derive a lower bound for the variance of the Mann-Whiteny Statistic.

If $\varphi(x, y) = \psi(y - x)$, φ is strictly convex in y and it is easy to verify that $(\partial^2\varphi)/(\partial x \partial y) < 0$. We know that the solution exists and is unique, and the problem is reduced to that of finding an η_1 and η_2 so that F_0 solves $P_{\mathfrak{F}\varphi\eta_1\eta_2}$ where $P_{\mathfrak{F}\varphi\eta_1\eta_2}$ is the problem of finding $F \in \mathfrak{F}$ which minimizes

$$\int_{-X}^{X} [\psi'(F_0(x) - x) + \eta_1 + \eta_2 x]F(x)\,dx.$$

Then by the theorems of Section 5, we know that $F_0(x)$ is given in terms of $g_{\eta_1\eta_2}(x)$, where $g_{\eta_1\eta_2}(x)$ is uniquely expressed in terms of the function

$$f_{\eta_1\eta_2}(x) = x + \psi'^{-1}(-\eta_1 - \eta_2 x).$$

Returning to our example, we have the function

$$f_{\eta_1\eta_2}(x) = (1 - \eta_2)x - \eta_1,$$

$$f_{\eta_1\eta_2}(x) = 0 \quad \text{for} \quad x_1 = \frac{\eta_1}{1 - \eta_2}, \qquad f_{\eta_1\eta_2}(x) = 1 \quad \text{for} \quad x_2 = \frac{1 + \eta_1}{1 - \eta_2}.$$

Case 1. $\eta_2 < 1$. Then $f_{\eta_1\eta_2}$ is increasing. Define $g_{\eta_1\eta_2}(x)$ by the following.

$$g_{\eta_1\eta_2}(x) = \begin{cases} 0, & x < \max(-X, x_1), \\ 1, & x \geqq \min(x_2, X), \\ (1 - \eta_2)x - \eta_1, & \text{elsewhere.} \end{cases}$$

As $g_{\eta_1\eta_2}$ is a cdf, $F_0(x) = g_{\eta_1\eta_2}(x)$.

Case 2. $\eta_2 \geqq 1$, $f_{\eta_1\eta_2}(x)$ is nonincreasing, and hence the solution is either a one-point distribution or a two-point distribution with total probability concentrated at $-X$ and X. In both cases, then, the solution is not admissible.

EXAMPLE 2. Consider the same problem as in Example 1, but with an additional restriction on the cdf $F(x)$, i.e., $F(x) \geqq x$. Now let $F(x)$ be a cdf on [0, 1].

Under this additional restriction, the class $\mathfrak{A}^*$ of admissible cdf's is also compact and convex. Then the solution to this problem exists. It is unique since $\varphi(x, y) = \frac{1}{2}(y - x)^2$ is strictly convex in y.

Applying the methods used to prove Lemma 4.2, we see that the problem is the same as that of minimizing

$$\int_0^1 [F_0(x) - x]F(x)\,dx$$

over all $F \in \mathfrak{A}^*$. It is easy to see that the set analogous to Γ_2 of Lemma 4.3 here is also convex, closed and bounded, and hence, applying the method of Lemma 4.4, we reduce the problem to that of finding the cdf's corresponding to

$$\min_{F \in \mathfrak{F}} \int_0^1 [F_0(x) - x + \eta_1 + \eta_2 x]F(x)\,dx$$

or

$$\min_{F \in \mathfrak{F}} \int_0^1 [F_0(x) + \eta_1 + \eta_3 x]F(x)\,dx, \qquad \eta_3 = \eta_2 - 1$$

where $\mathfrak{F}$ is the class of all cdf's F on [0, 1] such that $F(x) \geqq x$. We can now apply the methods of Section 5. Define the function $f_{\eta_1\eta_2}$ with $x \leqq f_{\eta_1\eta_2}(x) \leqq 1$ such that

$$f_{\eta_1\eta_2}(x) = -\eta_1 - \eta_2 x.$$

Define the function $g_{\eta_1\eta_2}$ such that

$$g_{\eta_1\eta_2}(x) = \begin{cases} f_{\eta_1\eta_2}(x), & \text{where } f_{\eta_1\eta_2} \text{ is defined,} \\ x \text{ or } 1, & \text{elsewhere on } [0, 1], \end{cases}$$

$g_{\eta_1\eta_2}(x)$ is continuous on $[0, 1]$, and $g_{\eta_1\eta_2}(1) = 1$.

Then $g_{\eta_1\eta_2}$ gives the solution F_0 of the problem if $g_{\eta_1\eta_2}(x)$ is a cdf. We shall give an explicit characterization of $g_{\eta_1\eta_2}$ in the various possible cases. Let the point of intersection of $y = -\eta_1 - \eta_2 x$ and $y = x$ be denoted by $x^* = -[\eta_1/(1 + \eta_2)]$. Let x^{**} be such that $-\eta_1 - \eta_2 x^{**} = 1$.

Case I. $\eta_2 > -1$.

(a) $x^* \leqq 0$

$$g_{\eta_1\eta_2}(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leqq x < 1. \\ 1, & x \geqq 1. \end{cases}$$

This $g_{\eta_1\eta_2}$ is a cdf, but it is not admissible.

(b) $x^* > 1$

$$g_{\eta_1\eta_2}(x) = \begin{cases} 0, & x < 0, \\ -\eta_1 - \eta_2 x, & 0 < x < x^{**}, \\ 1, & x > x^{**}. \end{cases}$$

If $x^{**} > 0$, $g_{\eta_1\eta_2}$ is a cdf and hence $F_0(x) = g_{\eta_1\eta_2}(x)$.

If $x^{**} < 0$, the solution is a one-point distribution with mass concentrated at $x = 0$, which is not admissible.

(c) $0 < x^* < 1$. Consider (i) $-1 < \eta_2 < 0$

$$g_{\eta_1\eta_2}(x) = \begin{cases} 0, & x < 0 \\ -\eta_1 - \eta_2 x, & 0 \leqq x < x^* \\ x, & x^* \leqq x \leqq 1 \\ 1, & x \geqq 1. \end{cases}$$

This is again a cdf, and hence $F_0(x) = g_{\eta_1\eta_2}(x)$.

(ii) $\eta_2 > 0$

$$g_{\eta_1\eta_2}(x) = \begin{cases} 0, & x < 0, \\ -\eta_1 - \eta_2 x, & 0 \leqq x \leqq x^*, \\ x, & x^* \leqq x \leqq 1, \\ 1, & x \geqq 1. \end{cases}$$

This $g_{\eta_1\eta_2}$ is not a cdf. Here $A(x) = F_0(x) + \eta_1 + \eta_2 x$. At $x = 0$, $A(x) > 0$ if $F_0(x) > -\eta_1$ and $A(x) < 0$ if $F_0(x) < -\eta_1$. Suppose there is a jump at $x = 0$ such that $F_0(0) > -\eta_1$, then $A(0) > 0$, and $F_0(x)$ is then continuous at $x = 0$. Hence there is a contradiction. Let there be a jump at $x = 0$ such that $c = F_0(0) < -\eta_1$, $A(0) < 0$ and hence we take as a possible minimizing cdf

$$G(x) = \begin{cases} 0, & x < 0, \\ c, & 0 \leqq x < c, \\ x, & c \leqq x < 1, \\ 1, & x > 1. \end{cases}$$

The remark after Theorem 5.1 puts the following restriction on c,

$$\int_0^c (c + \eta_1 + \eta_2 x)\, dx = 0,$$

i.e.,

$$(c + \eta_1)c + \frac{\eta_2}{2} c^2 = 0, \qquad \text{or} \qquad c = -[\eta_1/(\eta_2/2 + 1)],$$

since $c = 0$ gives an inadmissible cdf. Incidentally this shows also, as is evident from the value of c itself that

$$x^* < c < -\eta_1 .$$

The unique value of c which is obtained from the constraints exists if and only if

$$(1 - 2\mu_1)^3 = (1 - 3\mu_2)^2. \tag{7.1}$$

This condition is obtained by eliminating c between the equations

$$\int_0^1 F(x)\, dx = 1 - \mu_1 = c^2 + \frac{1 - c^2}{2} = \frac{c^2 + 1}{2}$$

and

$$2\int_0^1 xF(x)\, dx = 1 - \mu_2 = c^3 + \frac{2}{3}(1 - c^3) = \frac{c^3 + 2}{3}.$$

Hence $G(x)$ is admissible and $F_0(x) = G(x)$ if and only if μ_1 and μ_2 are such that (7.1) is satisfied.

Case II. $\eta_2 < -1$.

(a) $x^* < 0$. We then have

$$g_{\eta_1\eta_2}(x) = \begin{cases} 0, & x < 0, \\ -\eta_1 - \eta_2 x, & 0 \leqq x < x^{**}, \\ 1, & x \geqq x^{**}. \end{cases}$$

If $x^{**} > 0$, $g_{\eta_1\eta_2}$ is a cdf and hence $F_0(x) = g_{\eta_1\eta_2}(x)$. If $x^{**} < 0$, the minimizing cdf is a one-point distribution and is not admissible.

(b) $x^* > 1$. The minimizing cdf is the same as in Case I (a) and is not admissible.

(c) $0 < x^* < 1$. Then we have

$$g_{\eta_1\eta_2}(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leqq x \leqq x^*, \\ -\eta_1 - \eta_2 x, & x^* < x < x^{**}, \\ x, & x > x^{**}. \end{cases}$$

$g_{\eta_1\eta_2}$ is a cdf, and hence $F_0(x) = g_{\eta_1\eta_2}(x)$.

EXAMPLE 3. Let $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ be n ordered, independent observations from a cdf $F(x)$. Consider the problem of maximizing $E(x_n)$ with restrictions (3.2). The same problem for cdf's defined over the whole real line with restrictions on mean and variance has been discussed by Gumbel [10] and David and Hartley [1].

$$E(x_n) = \int_{-X}^{X} x\, d\{F(x)\}^n.$$

Integrating by parts, the above problem reduces to that of finding

$$\min_{F \in \Omega} \int_{-X}^{X} F^n(x)\, dx.$$

Now as $\varphi(x, y) = y^n$ is strictly convex in y and is a function of y alone, the solution $F_0(x)$ is given by the function $g_{\eta_1\eta_2}(x)$, where

$$g_{\eta_1\eta_2}(x) = \begin{cases} 0, & x < \max\left(-\dfrac{\eta_1}{\eta_2}, -X\right), \\ 1, & x \geqq \min\left(X, -\dfrac{n+\eta_1}{\eta_2}\right); \\ \left(-\dfrac{\eta_1 + \eta_2 x}{n}\right)^{1/n-1}, & \text{elsewhere.} \end{cases}$$

Here η_1 and η_2 are determined by the following four cases:

	$\max(x_1, -X)$	$\min(x_2, X)$
Case 1.	x_1	x_2
Case 2.	$-X$	x_2
Case 3.	$-X$	X
Case 4.	x_1	X

We give below the equations determining η_1, η_2 in the above cases.

Case 1. $\mu_1 = -\dfrac{1}{\eta_2}(1 + \eta_1)$,

$$\mu_2 = \frac{1}{\eta_2^2}\left(\eta_1^2 + 2\eta_1 + \frac{n^2}{2n-1}\right).$$

Case 2. $\mu_1 = -\dfrac{n-1}{\eta_2}\xi^{n/(n-1)} - \dfrac{1+\eta_1}{\eta_2}, \qquad \xi = \dfrac{1}{n}(-\eta_1 + \eta_2 X),$

$$\mu_2 = \left(\frac{n+\eta_1}{\eta_2}\right)^2 - \frac{2}{\eta_2^2}\left[\frac{n^2(n-1)}{2n-1} + (n-1)\eta_1 - (n-1)\xi^{n/(n-1)} \times \left(\eta_1 + \frac{n^2}{n-1}\xi\right)\right].$$

Case 3. $\mu_1 = X + \frac{n-1}{\eta_2}(\zeta^{n/n-1} - \xi^{n/n-1}), \qquad \zeta = \frac{1}{n}(-\eta_1 - \eta_2 X),$

$$\mu_2 = X^2 - \frac{2n}{\eta_2}\left[\frac{n-1}{n}\eta_1(\zeta^{n/n-1} - \xi^{n/n-1}) + \frac{n(n-1)}{2n-1} \cdot (\zeta^{(2n-1)/(n-1)} - \xi^{(2n-1)/(n-1)})\right].$$

Case 4. $\mu_1 = X + \frac{n-1}{\eta_2}\zeta^{n/(n-1)},$

$$\mu_2 = X^2 - \frac{n}{\eta_2}\left[\frac{(n-1)\eta_1}{n}\zeta^{n/(n-1)} + \frac{n(n-1)}{2n-1}\xi^{(2n-1)/(n-1)}\right].$$

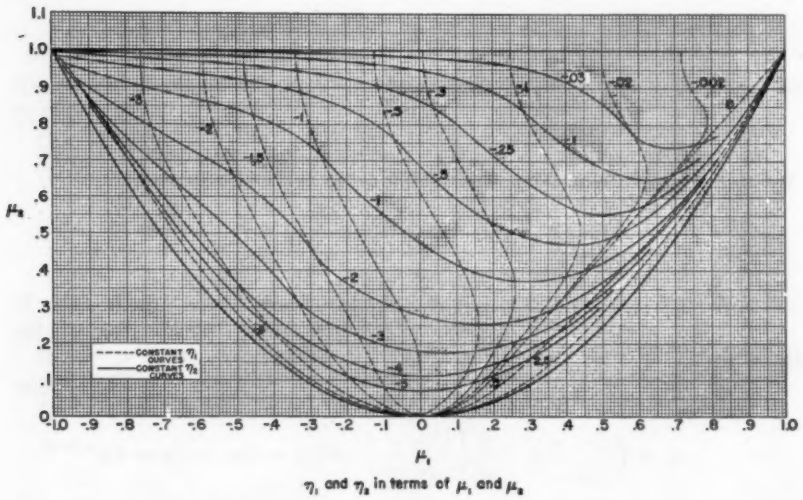
Fig. 2 represents η_1 and η_2 in terms of μ_1 and μ_2 for the case $X = 1$.

Similar results can be easily obtained for maximizing the expectations of the range and the smallest observation with similar restrictions on the underlying cdf's.

8. Characterization of the solution of the maximum problem and some examples. In this section we find the solution to the problem of maximizing

$$I(F) = \int_{-X}^{X} \varphi(x, F(x))\, dx$$

over all admissible cdf's. The existence of the maximizing cdf has already been established by Lemma 3.2. We shall show now that the solution is a discrete distribution and in our case, is at most a three-point distribution, i.e., a distribu



η_1 and η_2 in terms of μ_1 and μ_2

FIG. 2

tion concentrating all its mass at just three or fewer points. Some illustrations have been given at the end of this section.

THEOREM 8.1. *The solution to the problem of maximizing $I(F)$ over the class $\mathfrak{A}$ of admissible cdf's, is at most a three-point distribution.*

PROOF. The inequality (4.0) shows that a convex combination of two maximizing admissible cdf's which is itself also admissible, gives a value which is smaller than the maximum. Hence the maximum of $I(F)$ occurs for cdf's which correspond to the extreme points of the convex set $\mathfrak{A}$.

By Theorems 21.1 and 21.3 of Karlin and Shapley [5], it is then easy to see that the maximizing cdf is at most a three-point distribution.

REMARK. It is important to note here that in some cases, the maximizing admissible cdf can be further reduced to a two-point distribution [1], [3].

We shall illustrate the above results by a few examples.

EXAMPLE 1. Suppose we want to minimize $E(x_n)$ given in Example 4 of the last section, over all admissible cdf's. The problem is the same as that of maximizing

$$\int_{-X}^{X} F^n(x)\,dx$$

over all admissible cdf's. As $\varphi(x, y) = y^n$, is strictly convex in y, the maximizing admissible cdf of the above integral is at most a three-point distribution.

Similarly the minimizing admissible cdf of $E(w_n)$ is at most a three-point distribution. David and Hartley [1] claim that it can be further reduced to a two-point distribution.

EXAMPLE 2. Suppose we are interested in finding the maximizing cdf of

$$\frac{1}{2}\int_{-X}^{X} [F(x) - x]^2\,dx$$

such that $F(x)$ satisfies side conditions (3.3).

Now $\varphi(x, y) = \frac{1}{2}(y - x)^2$ is strictly convex function of y. Hence the solution of the above problem is at most a three-point distribution.

9. Acknowledgment. The author is deeply grateful to Professor Herman Chernoff who suggested the problem and gave generous help and guidance throughout the entire work.

REFERENCES

[1] H. O. HARTLEY AND H. A. DAVID, "Universal bounds for mean range and extreme observations," *Ann. Math. Stat.*, Vol. 25 (March, 1954), pp. 85–89.

[2] H. CHERNOFF AND S. REITER, "Selection of a Distribution Function to Minimize an Expectation Subject to Side Conditions," Technical Report No. 23 (March 12, 1954), Applied Math. and Statistics Laboratory, Stanford University.

[3] S. KARLIN, "Notes on Theory of Games," California Institute of Technology, mimeo. (1954).

[4] S. KARLIN, "Notes on Theory of Games and Decision Functions," Stanford University, mimeo. (1955).

[5] S. Karlin and L. S. Shapley, "Geometry of Moment Spaces," *Memoirs of American Mathematical Society*, No. 12.

[6] H. D. Brunk, G. M. Ewing and W. R. Utz, "Minimizing Integrals in Certain Classes of Monotone Functions with Applications," University of Missouri, Department of Mathematics, September, 1954.

[7] Z. W. Birnbaum and O. M. Klose, "Bounds for the variance of the U-statistic in terms of probability $Y < X$," Abstract, *Bull. Amer. Math. Soc.*, 1955.

[8] M. Loève, "Probability Theory," D. Van Nostrand, New York, 1955.

[9] R. L. Plackett, "Limits of the ratio of mean range to standard deviation," *Biometrika*, Vol. 34 (1947), pp. 120–122.

[10] E. J. Gumbel, "The maxima of the mean largest value and of the range," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 76–84.

[11] W. Hoeffding, "The extrema of the expected value of a function of independent random variables," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 268–275.

[12] S. H. Khamis, "On the reduced moment problem," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 113–122.

[13] H. Rubin and S. Isaacson, "On Minimizing an Expectation Subject to Certain Side Conditions," Technical Report No. 25 (July, 1954).

[14] S. Moriguti, "Extremal properties of extreme value distributions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 523–536.

[15] W. Hoeffding and S. S. Shrikhande, "Bounds for the distribution function of a sum of independent identically distributed random variables," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 439–449.

ON CONSISTENT ESTIMATES OF THE SPECTRUM OF A STATIONARY TIME SERIES[1, 2]

By Emanuel Parzen[3]

Columbia University, Hudson Laboratories, Dobbs Ferry, N. Y.

Summary. This paper is concerned with the spectral analysis of wide sense stationary time series which possess a spectral density function and whose fourth moment functions satisfy an integrability condition (which includes Gaussian processes). Consistent estimates are obtained for the spectral density function as well as for the spectral distribution function and a general class of spectral averages. Optimum consistent estimates are chosen on the basis of criteria involving the notions of order of consistency and asymptotic variance. The problem of interpolating the estimated spectral density, so that only a finite number of quantities need be computed to determine the entire graph, is also discussed. Both continuous and discrete time series are treated.

1. Introduction. A stochastic process is a family of random variables $x(t)$, where t varies in some set T. If the set T is the infinite real line, then $x(t)$ is called a random function, and if $T = \{0, \pm 1, \pm 2, \cdots\}$, then $x(t)$ is called a random sequence. If the parameter t is interpreted as denoting time, then the stochastic process is called a time series, with the adjectives continuous or discrete being used to indicate whether it is a random function or a random sequence.

Let us suppose that we have observed a sample of length T of a (continuous or discrete) time series $x(t)$. The general problem of time series analysis is to infer the statistical characteristics of $x(t)$ from the observed sample. Now in order to perform a statistical analysis of $x(t)$, one has to assume a model for $x(t)$ which is completely specified except for the values of certain parameters which one proceeds to estimate on the basis of the observed sample.

A widely adopted model for $x(t)$ (see Grenander and Rosenblatt [4], [5]) is the following. It is assumed that $x(t)$ may be written as a sum of a mean value function $m(t)$ and a fluctuation function $y(t)$:

$$x(t) = m(t) + y(t). \tag{1.1}$$

The domain T of the variable t is to be taken as the infinite real line, $-\infty < t < \infty$, in the continuous case, and as the set of integers $0, \pm 1, \pm 2, \cdots$ in the discrete case. We seek to treat simultaneously both discrete and continuous time

Received September 13, 1955; revised October 15, 1956.

[1] This work was supported by the Office of Naval Research under Contract N6-ONR-27135. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Hudson Laboratories Contribution No. 15. A more detailed version of this paper may be found in Hudson Laboratories Technical Report No. 33.

[3] Now at Stanford University.

series. Most equations will hold for both cases, with the proper interpretation, which will be explained as we proceed.

It is assumed that the function $m(t)$ is nonrandom, and that there is a fixed number K of known functions $\varphi_1(t), \cdots, \varphi_K(t)$ such that $m(t)$ may be written as a linear combination of the $\varphi_j(t)$:

$$m(t) = m_1\varphi_1(t) + \cdots + m_K\varphi_K(t). \tag{1.2}$$

The constants m_j (for $j = 1, \cdots, K$) are unknown, and are to be estimated from the sample.

The fluctuation function $y(t)$ is a stochastic process, whose mean value function $Ey(t)$ vanishes identically in t. It is assumed that it possesses a finite second moment $E \mid y(t) \mid^2$, and that it is wide sense stationary, which means that the product moment $Ey(t)\ y(t + v)$ is independent of t, and depends only on v. One then defines the covariance function

$$R(v) = Ey(t)\ y(t + v). \tag{1.3}$$

In the case of random functions, it is assumed that $R(v)$ is continuous. Then, $R(v)$ possesses a representation as a Fourier-Stieltjes integral:

$$R(v) = \int e^{iwv}\, dF(w), \tag{1.4}$$

where $F(w)$ is a bounded non-decreasing function, called the spectral distribution function of the process. The domain of the variable v is the same as that of t, and the domain of the variable w is $-\infty$ to ∞ in the continuous case, and $-\pi$ to π in the discrete case. The domain of integration of an integral involving w is to be taken as the whole domain of w, in cases where it is not otherwise specified.

It is assumed next that $R(v)$ is summable. It then follows that the spectral distribution function $F(w)$ possesses a continuous density function $f(w)$, called the spectral density function of the time series $x(t)$. The following relations hold:

$$R(v) = \int e^{iwv} f(w)\, dw, \tag{1.5}$$

$$f(w) = \frac{1}{2\pi} \int e^{-iwv} R(v)\, dv \tag{1.6c}$$

$$= \frac{1}{2\pi} \sum e^{-iwv} R(v). \tag{1.6d}$$

In cases where the limits of integration (or summation) of an integral (or sum) involving the variables u or v are omitted, they are to be assumed to be $-\infty$ to $+\infty$. Henceforth, we write equations of the type of (1.6) only once, for the continuous case, with the understanding that for every such equation a corresponding equation may be written for the discrete case by replacing the integral by a sum. For certain important equations, we will write, without further explanation, two equations, with a suffix d for the discrete case and a suffix c for the continuous case.

The model for the process $x(t)$ which has just been described assumes only a knowledge of the first and second moments of the process, and assumes no knowledge of the probability distribution. The moments are assumed to be completely specified by the constants $m_1, \cdots, m_K$, and the covariance function $R(v)$, or equivalently the spectral density function $f(w)$. By analysis of an observed time series is meant the estimation of the value of these quantities on the basis of observed samples. The estimation of the constants $m_1, \cdots, m_K$ is called regression analysis, and the estimation of the spectral functions is called spectral analysis.

A basic requirement for an estimate is that it be *consistent* in quadratic mean. Let m be an unknown parameter of a time series $x(t)$, and let $x(t)$ for $0 \leqq t \leqq T$ (or $t = 1, \cdots, T$ in the discrete case) be an observed sample of the time series. An estimate m_T of m, formed on the basis of the sample, is said to be *consistent* in quadratic mean if the mean square error $E \mid m_T - m \mid^2$ tends to zero, as $T \to \infty$, where the expected value is taken under the assumption that m is the true parameter value. If an estimate is consistent, it is then asymptotically *unbiased*, which means that $Em_T \to m$ as $T \to \infty$.

However, we shall be interested in estimates which are consistent and asymptotically unbiased at certain prescribed rates. Let α be a positive number. We define an estimate to be asymptotically unbiased of the order of T^α if, for some finite constant β,

$$\lim_{T \to \infty} T^\alpha (Em_T - m) = \beta. \tag{1.7}$$

We say that an estimate possesses an asymptotic variance σ^2 of the order of $T^{2\alpha}$ if σ is positive and

$$\lim_{T \to \infty} T^\alpha \sigma[m_T] = \sigma, \tag{1.8}$$

where $\sigma^2[m_T] = E \mid m_T - Em_T \mid^2$ is the variance of m_T. The importance of these notions derives from the central limit theorem, for dependent random variables, from which one may hope to obtain conditions that the normalized random variable $(m_T - Em_T)/\sigma[m_T]$ tends to a normal distribution. We define an estimate to be consistent of the order of $T^{2\alpha}$, with asymptotic bias β and asymptotic variance σ^2, if (1.7) and (1.8) hold. If such an estimate obeys the Central Limit Theorem, then the random variable $T^\alpha (m_T - m)$ tends to a normal distribution with mean β and variance σ^2. Many estimates that one encounters are consistent of the order of T; however, we will encounter below estimates which are consistent of a lower order.

A knowledge of the order of consistency, the asymptotic bias, and the asymptotic variance of an estimate is valuable on several counts, as will be shown in detail in a later paper [8].

The problem of regression analysis has been extensively treated by Grenander and Rosenblatt in several excellent papers (see [5]), in which they obtained expressions for the asymptotic variances (of order T) of various estimates of the constants m_j in the model given above, and obtained conditions that the least

squares estimate and the best linear unbiased estimates have the same asymptotic variance. We mention regression analysis here only to point out that the results of this paper remain valid if in estimating the spectrum one uses the deviations of the observed values of $x(t)$ from the sample mean value function formed by inserting into (1.2) the least squares estimates of the constants m_j. As far as detailed considerations are concerned, we consider only the case where $m(t) = m$, an unknown constant.

The problem of outstanding interest at the present time in the analysis of time series is that of estimating the spectral density function, and it is this problem that is treated in this paper. In view of Eq. (1.6), the obvious way to estimate $f(w)$ is to form the Fourier transform $f_T(w)$ of the least squares estimate $R_T(v)$ of the covariance. The sample spectral density function $f_T(w)$ so obtained is essentially what has been studied by various authors under the name of the *periodogram*. However, as is well known, it turns out that $f_T(w)$ is not a consistent estimate of $f(w)$.

Rather, to begin with, we are only able to estimate what may be called *spectral averages*; that is, averages of the spectral density function of the form

$$J(A) = \int A(w)f(w)\,dw, \tag{1.9}$$

where $A(w)$ is a suitably chosen function. On the one hand $A(w)$ may be chosen to be a unit step function, $A(w) = 1$ or 0 according as $w < w_0$ or $w \geqq w_0$. Then $J(A)$ represents the spectral distribution function $f(w_0)$. On the other hand, $A(w)$ may be a function highly peaked about a center frequency w_0.

In Section 5, we obtain a class of consistent estimates of the spectral density function at a point w_0. However, the order of consistency of these estimates will be $T^{2\alpha}$, where $0 < \alpha < \frac{1}{2}$. Expressions are obtained for the asymptotic variance and bias of such estimates, so that the means are at hand for choosing among the large class of estimates presented. In Section 6, consistent estimates of the spectral density function, asymptotically optimum within the family of estimates considered, are discussed. In Section 7, we use the ideas leading to consistent estimates of the spectral density to obtain alternative estimates of the spectral averages. In Section 8, we treat the problem of interpolating the spectral density.

2. Assumptions on the fourth moments. Some additional assumptions are required in addition to the assumptions we have already stated. We assume that the fluctuation function $y(t)$ is wide sense stationary of order 4, in the sense that $E\,|\,y(t)\,|^4$ exists for all t, and the fourth moment function

$$P(v_1, v_2, v_3) = Ey(t)\,y(t + v_1)\,y(t + v_2)\,y(t + v_3) \tag{2.1}$$

is a function only of the time differences v_1, v_2, v_3, and not of the initial time t.

Now if the process $y(t)$ were normally distributed, then $P(v_1, v_2, v_3)$ could be expressed in terms of the covariance function $R(v)$ as follows:

$$P_G(v_1, v_2, v_3) = R(v_1)R(v_2 - v_3) + R(v_2)R(v_3 - v_1) + R(v_3)R(v_1 - v_2). \tag{2.2}$$

We introduce the function

$$Q(v_1, v_2, v_3) = P(v_1, v_2, v_3) - P_G(v_1, v_2, v_3), \tag{2.3}$$

which is the difference between the actual fourth moment function of $y(t)$, and what it would be if $y(t)$ were Gaussian. We refer to $Q(v_1, v_2, v_3)$ as the non-Gaussian part of the fourth moment function of $y(t)$; it is the same as the fourth cumulant function.

We assume that $Q(v_1, v_2, v_3)$ is absolutely summable (and, in the continuous parameter case, continuous) over all of (v_1, v_2, v_3) space.

We will find in many instances that $Q(v_1, v_2, v_3)$ admits of a representation as a Fourier integral:

$$Q(v_1, v_2, v_3) = \iiint \exp[i(w_1 v_1 + w_2 v_2 + w_3 v_3)] g(w_1, w_2, w_3)\, dw_1 dw_2 dw_3, \tag{2.4}$$

where the function $g(w_1, w_2, w_3)$ is absolutely integrable over all of (w_1, w_2, w_3) space. We may have also the relation

$$\int du Q(v_1, u, u + v_2) = 2\pi \iint dw_1 dw_2 g(w_1, -w_2, w_2) \exp[i(w_1 v_1 + w_2 v_2)]. \tag{2.5}$$

We will assume these relations to be valid, since they simplify the writing of some of the results. It should be pointed out that the notion of the Fourier transform of the non-Gaussian part of the fourth moment function has previously been considered by Magness [6] where some examples may be found.

In the continuous parameter case we assume also that the stochastic process $x(t, \omega)$, where ω varies in a space Ω on which the basic probability measure P is defined, is measurable jointly in t and ω. Then the random integrals, such as $\int_0^T x(t)\, dt$, which are employed exist with probability one, by virtue of the Fubini theorem (see Doob [9]). Alternatively, the random integrals employed may be defined as limits in quadratic mean (see Loève [10]).

3. The sample covariance and spectral density functions. The estimates of the spectrum that we shall consider will be defined in terms of two functions, the sample covariance function and the sample spectral density function, which are defined in this section. Given a sample of observed values of $x(t)$ for $0 \leqq t \leqq T$ (or for $t = 1, \cdots, T$), let m_T be the least squares estimate of m, and consider the function $Y_T(t)$, defined by

$$\begin{aligned} Y_T(t) &= x(t) - m_T \qquad \text{for } 0 \leqq t \leqq T, \\ &= 0 \qquad\qquad\quad \text{otherwise.} \end{aligned} \tag{3.1}$$

Define now the function

$$f_T(w) = \frac{1}{T}\left| \int_0^T Y_T(t) e^{-itw}\, dt \right|^2 \tag{3.2c}$$

$$= \frac{1}{T}\left| \sum_{t=1}^{T} Y_T(t) e^{-itw} \right|^2, \tag{3.2d}$$

which may be regarded as the notion of the "periodogram" extended to the case of time series with an unknown mean value. We call $f_T(w)$ the sample spectral density function, because its Fourier integral

$$R_T(v) = \int e^{ivw} f_T(w)\, dw \tag{3.3}$$

is a consistent estimate of the covariance function. We call $R_T(v)$ the sample covariance function. It vanishes for $|v| \geqq T$, and for $|v| < T$,

$$R_T(v) = \frac{1}{T}\int_0^{T-|v|} Y_T(t)Y_T(t + |v|)\, dt \tag{3.4c}$$

$$= \frac{1}{T}\sum_{t=1}^{T-|v|} Y_T(t)Y_T(t + |v|). \tag{3.4d}$$

We may invert (3.3) to obtain

$$f_T(w) = \frac{1}{2\pi}\int_{-T}^{T} e^{-ivw} R_T(v)\, dv \tag{3.5c}$$

$$= \frac{1}{2\pi}\sum_{|v|\leq T} e^{-ivw} R_T(v). \tag{3.5d}$$

In the continuous parameter case, the interval of integration in (3.3) is infinite, and to establish that $f_T(w)$ is summable, one needs to employ a standard argument involving Plancherel's theorem.

An important role in the sequel will be played by the following representation of $R_T(v)$, for $|v| \leqq T$:

$$R_T(v) = D_T(v) + b_T(v) + R(v)\left(1 - \frac{|v|}{T}\right), \tag{3.6}$$

where

$$D_T(v) = \frac{1}{T}\int_0^{T-|v|} dt\{y(t)y(t + |v|) - R(v)\} \tag{3.7}$$

and $b_T(v)$ is defined so as to make Eq. (3.6) correct.

The term $b_T(v)$ represents the bias arising from the fact that the sample covariances are computed using $Y_T(t)$, the deviations of the observations from the sample mean. That it may be essentially ignored in our calculations will follow from the fact that there is a constant K such that

$$T^2 E|b_T(v)|^2 \leqq K^2 \tag{3.8}$$

for any v and T. To establish (3.8), it suffices to show that there is a constant K' such that, for any choice of numbers T and T_1, T_2, T_3, T_4 satisfying $0 \leqq T_1, T_2, T_3, T_4 \leqq T$,

$$E\left|\int_{T_1}^{T_2}\int_{T_3}^{T_4} y(t_1)y(t_2)\, dt_1\, dt_2\right|^2 \leqq K'T^2, \tag{3.9}$$

which follows from the fact that the expected value in (3.9) is less than

$$3\left\{T\int |R(u)|\,du\right\}^2 + T\iiint |Q(u_1, u_2, u_3)|\,du_1\,du_2\,du_3.$$

We next evaluate the covariance of $D_T(v)$. We obtain that, for any non-negative numbers v_1 and v_2,

$$TED_T(v_1)D_T(v_2) = \int_{-T}^{T} du U_T(u, v_1, v_2)\{Q(v_1, u, u+v_2) + R(u)R(u+v_1-v_2) + R(u+v_1)R(u-v_2)\}, \tag{3.10}$$

where $U_T(u, v_1, v_2)$ is a function with values between 0 and 1 defined as follows:

$$\begin{aligned} U_T(u, v_1, v_2) &= 0 && u \leqq -T + v^2 \\ &= 1 - \frac{v_2 + u}{T} && -T + v_2 \leqq u \leqq \min(0, v_2 - v_1) \\ &= 1 - \frac{\max(v_1, v_2)}{T} && \min(0, v_2 - v_1) \leqq u \leqq \max(0, v_2 - v_1) \\ &= 1 - \frac{v_1 + u}{T} && \max(0, v_2 - v_1) \leqq u \leqq T - v_1 \\ &= 0 && T - v_1 \leqq u \end{aligned} \tag{3.11}$$

To establish (3.10) one makes the change of variable $u = t_2 - t_1$, $v = t_2$ in the expression

$$T^2 ED_T(v_1)D_T(v_2) = \int_0^{T-v_1}\int_0^{T-v_2} dt_1\,dt_2 \{Q(v_1, t_2 - t_1, t_2 - t_1 + v_2) + R(t_2 - t_1)R(t_2 - t_1 + v_2 - v_1) + R(t_2 - t_1 + v_2)R(t_2 - t_1 - v_1)\}.$$

As a first consequence of (3.8) and (3.10), we obtain the following theorem.

THEOREM 3: *For any non-negative numbers* v, v_1, *and* v_2,

$$\lim_{T\to\infty} T^{1/2}\,|ER_T(v) - R(v)| = 0, \tag{3.12}$$

$$\begin{aligned} &\lim_{T\to\infty} T\,\mathrm{Cov}[R_T(v_1), R_T(v_2)] \\ &= \int du\,\{Q(v_1, u, u+v_2) + R(u)R(u+v_1-v_2) + R(u+v_1)R(u-v_2)\} \end{aligned} \tag{3.13}$$

$$\begin{aligned} &= 2\pi\left\{\iint dw_1\,dw_2 \exp[i(w_1 v_1 + w_2 v_2)]g(w_1, -w_2, w_2)\right. \\ &\qquad \left. + \int dw f^2(w)\exp[iw(v_1 - v_2)] + \int dw f^2(w)\exp[iw(v_1 + v_2)]\right\}. \end{aligned} \tag{3.14}$$

4. Estimation of spectral averages. To estimate the spectral average $J(A)$ there are two methods available, which may be called the method of filtering and the method of covariance averages. In the method of filtering, one estimates the variance (zero lag covariance) of a new time series obtained by filtering the observed series. In the method of covariance averages, one defines a sample spectral average $J_T(A)$, which may be expressed as an average with respect to the sample spectral density function or with respect to the sample covariances. This latter form is the more convenient for computations. Only the method of covariance averages is discussed here.

Spectral averaging functions: A function $A(w)$ will be called a spectral averaging function if it is a real valued function which is both absolutely integrable and square integrable. Its Fourier transform

$$a(v) = \frac{1}{2\pi}\int e^{-ivw}A(w)\,dw \tag{4.1}$$

is then bounded and square integrable. We call $a(v)$ a covariance averaging function. We assume finally that

$$|\,a(v)\,| = o(|\,v\,|^{-r}) \qquad \text{for some } r > \tfrac{1}{2}. \tag{4.2}$$

If $A(w)$ has finite total variation (and also, in the continuous parameter case, vanishes at infinity), then $|\,a(v)\,| = O(|\,v\,|^{-1})$. From (4.2) it follows that, for some constant K_1 and some $\epsilon > 0$,

$$\int_{-T}^{T} |\,a(v)\,|\,dv \leqq K_1 T^{1/2-\epsilon} \tag{4.3}$$

for all T, and also that

$$\int |\,v\,|^{1/2}\,|\,a(v)R(v)\,|\,dv < \infty. \tag{4.4}$$

A lemma: Of frequent use in the sequel will be the following lemma.

LEMMA 4: *Let $q \geqq 0$ and $s > 0$. Let M_T be a sequence of constants tending to ∞ as $T \to \infty$. Suppose that*

$$\int |\,v\,|^{q}\,|\,a(v)R(v)\,|\,dv < \infty. \tag{4.5}$$

Then, as $T \to \infty$,

$$M_T^q \int_{|v|\geqq M_T} |\,a(v)R(v)\,|\,dv \to 0, \tag{4.6}$$

$$\frac{1}{M_T^s}\int_{|v|\leq M_T} |\,v\,|^{q+s}\,|\,a(v)R(v)\,|\,dv \to 0. \tag{4.7}$$

Sample spectral averages: The spectral average $J(A)$ may be defined in terms of either the spectral density or the covariance function by

$$J(A) = \int A(w)f(w)\,dw = \int a(v)R(v)\,dv. \tag{4.8}$$

Accordingly, we define the sample spectral average $J_T(A)$ by

$$J_T(A) = \int A(w)f_T(w)\,dw = \int_{-T}^{T} a(v)R_T(v)\,dv. \tag{4.9}$$

The properties of $J_T(A)$ as an estimate of $J(A)$ are given in the following theorem:

THEOREM 4: *For any spectral averaging functions $A(w)$, $A_1(w)$, and $A_2(w)$,*

$$\lim_{T\to\infty} T^{1/2} \mid EJ_T(A) - J(A) \mid = 0, \tag{4.10}$$

$$\begin{aligned}\lim_{T\to\infty} T \operatorname{Cov}[J_T(A_1), J_T(A_2)] &= 4\pi \int dw f^2(w)A_1^0(w)A_2^0(w) \\ &\quad + 2\pi \iint dw_1\,dw_2\, g(w_1, -w_2, w_2)A_1^0(w_1)A_2^0(w_2),\end{aligned} \tag{4.11}$$

where

$$A^0(w) = \tfrac{1}{2}\{A(w) + A(-w)\}. \tag{4.12}$$

PROOF. Omitted, since it is similar to the proofs of Theorems 5A and 5B.

5. Estimation of the spectral density. Various authors have pointed out that the sample spectral density function, or periodogram, $f_T(w)$ is not a consistent estimate of the spectral density function $f(w)$. The suggestion has been made to estimate $f(w)$ at a point w_0 by averaging the values of $f_T(w)$ in the neighborhood of w_0. However, this yields a consistent estimate not of $f(w_0)$, but rather of the spectral average in the neighborhood of w_0. To eliminate this bias, one needs to narrow the neighborhood averaged over as the sample size is increased. The manner in which this is to be done is examined in this section. A similar method of obtaining a consistent estimate of the spectral density is that of Bartlett (see [1]), who has suggested a modified form of the periodogram. More general methods of constructing consistent estimates of the spectral density have been given by Grenander [3] and Tukey [7]. In this section these methods are generalized somewhat further. A noteworthy feature of the general method of constructing consistent estimates of the spectral density which is discussed in this section is that one may construct estimates which are consistent of any prescribed order $T^{2\alpha}$, where $0 < \alpha < \frac{1}{2}$.

Covariance averaging kernels: A function $k(z)$, defined for all real z, will be called a covariance averaging kernel if it fulfills the following conditions. It is even, bounded, square integrable, and normalized so that $k(0) = 1$. Its Fourier transform $K(w)$ is defined (as a limit in quadratic mean) to satisfy

$$k(z) = \int e^{-iwz}K(w)\,dw. \tag{5.1}$$

It is assumed that there is a constant K_1 and an $\epsilon > 0$ such that

$$B \int_{-T}^{T} | k(Bv) | \, dv \leqq K_1 (BT)^{1/2-\epsilon}, \tag{5.2c}$$

$$B \sum_{|v| \leq T} | k(Bv) | \leqq K_1 (BT)^{1/2-\epsilon}, \tag{5.2d}$$

for every B and T. A sufficient condition for (5.2) to hold is that $k(z)$ satisfy (4.2).

Given a kernel $k(z)$, and a positive number r, define

$$k^{(r)} = \lim_{z \to 0} \frac{1 - k(z)}{| z |^r}. \tag{5.3}$$

We assume next that there is a largest number r, called the *characteristic exponent* of the kernel $k(z)$, such that $k^{(r)}$ exists and is finite (nonzero). If the limit in (5.3) exists for every positive r, then the kernel is said to have characteristic exponent ∞.

Estimates of the spectral density: Let $k(z)$ be a covariance averaging kernel and let B_T be a sequence of constants tending to 0, as $T \to \infty$, in such a way that $B_T T \to \infty$. As an estimate of the spectral density function we define the even function

$$f_T^*(w) = \frac{1}{2\pi} \int_{-T}^{T} e^{-ivw} k(B_T v) R_T(v) \, dv \tag{5.4c}$$

$$= \frac{1}{2\pi} \sum_{|v| \leq T} e^{-ivw} k(B_T v) R_T(v). \tag{5.4d}$$

The constant B_T may be called the *bandwidth* of the estimate. In terms of the sample spectral density, one may write

$$f_T^*(w) = \frac{1}{B_T} \int_{-\infty}^{\infty} K\left(\frac{\lambda - w}{B_T}\right) f_T(\lambda) \, d\lambda, \tag{5.5}$$

where $f_T(\lambda)$ is to be defined as a periodic function in the discrete parameter case. Alternate ways in which $f_T^*(w)$ may be written in the discrete parameter case are

$$f_T^*(w) = \int_{-\pi}^{\pi} d\lambda f_T(\lambda) \frac{1}{B_T} \sum_{n=-\infty}^{\infty} K\left(\frac{\lambda - w - 2\pi n}{B_T}\right) \tag{5.5'}$$

and

$$f_T^*(w) = \int_{-\pi}^{\pi} d\lambda f_T(\lambda) K_T(\lambda - w), \tag{5.5''}$$

where $K_T(w)$ is defined so that

$$k(B_T v) = \int_{-\pi}^{\pi} e^{-ivw} K_T(w) \, dw.$$

Various estimates of the spectral density which have been proposed (see Bartlett [1], [2], Grenander [3], Tukey [7]) may be obtained as instances of (5.4).

By choosing $k(z) = 1 - |z|$ if $|z| \leqq 1$ and 0 otherwise and letting $B_T = (1/M)$, where M is an integer less than T, one has a modified form of Bartlett's estimate:

$$\frac{1}{2\pi} \sum_{v=-M}^{M} e^{-iwv} \left(1 - \frac{|v|}{M}\right) R_T(v).$$

By choosing $k(z) = \sin z/z$ and letting $B_T = h$, one has Daniell's estimate:

$$\frac{1}{2\pi} \sum_{v=-T}^{T} e^{-iwv} \frac{\sin(hv)}{hv} R_T(v) = \frac{1}{2h} \int_{-h}^{h} f_T(\lambda - w)\, d\lambda.$$

By choosing $k(z) = 1$ if $|z| < 1$ and 0 otherwise, and letting $B_T = (1/M)$, one has the truncated estimate

$$\frac{1}{2\pi} \sum_{v=-M}^{M} e^{-iwv} R_T(v),$$

which, in view of the fact that the Fourier transform $K(w)$ of $k(z)$ is not non-negative, has the possibly undesirable property that it is not necessarily non-negative.

The properties of the estimate $f_T^*(w)$ are embodied in the following theorems.

THEOREM 5A. *The asymptotic covariance of the estimate $f_T^*(w)$ defined by* (5.4) *satisfies, for any non-negative frequencies w_1 and w_2,*

$$\lim_{T\to\infty} TB_T \operatorname{Cov}[f_T^*(w_1), f_T^*(w_2)] = f^2(w) \int k^2(z)\, dz\{1 + \delta(0, w_1)\}\delta(w_1, w_2), \tag{5.6}$$

where $\delta(w_1, w_2) = 1$ if $w_1 = w_2$ and 0 if $w_1 \neq w_2$. Further, for any $\epsilon > 0$, the limit in (5.6) *is uniform in w_1 and w_2 such that $w_1 \geqq \epsilon$ and $w_2 \geqq \epsilon$.*

REMARK. The integral in (5.6) is *not* to be replaced by a summation in the discrete parameter case.

THEOREM 5B. *Let $q > 0$ be such that*

$$\int |v|^q |R(v)|\, dv < \infty, \tag{5.7c}$$

$$\sum |v|^q R(v) < \infty. \tag{5.7d}$$

Define the generalized qth *spectral derivative $f^{(q)}(w)$ by*

$$f^{(q)}(w) = \frac{1}{2\pi} \int e^{-iwv} |v|^q R(v)\, dv \tag{5.8c}$$

$$= \frac{1}{2\pi} \sum e^{-iwv} |v|^q R(v). \tag{5.8d}$$

Let the covariance averaging kernel $k(v)$ have characteristic exponent $r \geqq q$. Let the constants B_T be chosen so that

$$0 < \lim_{T\to\infty} TB_T^{1+2q} < \infty. \tag{5.9}$$

Then, uniformly in w,

$$\lim_{T\to\infty} B_T^{-2q} \mid Ef_T^*(w) - f(w) \mid = \mid k^{(q)} f^{(q)}(w) \mid^2 \qquad \text{if } r = q$$

$$= 0 \qquad \text{if } r > q. \tag{5.10}$$

PROOFS. We first show that the term $b_T(v)$, defined by (3.6), has no effect by showing that uniformly in w,

$$\lim_{T\to\infty} TB_T E \left| \int_{-T}^{T} dv e^{-ivw} k(B_T v) b_T(v) \right|^2 = 0. \tag{5.11}$$

By Minkowski's inequality, (3.8), and (5.2), the square root of the quantity in (5.11) whose limit is being taken is less than

$$K(TB_T)^{-1/2} B_T \int_{-T}^{T} dv \mid k(B_T v) \mid \leqq KK_1 T^{-\epsilon},$$

which tends to 0 as $T \to \infty$.

We next establish Theorem 5B. By (3.6), we write

$$2\pi f_T^*(w) = \int_{-T}^{T} dv e^{-ivw} k(B_T v) \left\{ D_T(v) + b_T(v) + R(v) \left(1 - \frac{\mid v \mid}{T} \right) \right\}. \tag{5.12}$$

Therefore

$$\begin{aligned} 2\pi B_T^{-q}(Ef_T^*(w) - f(w)) = {} & EB_T^{-q} \int_{-T}^{T} dv e^{-ivw} k(B_T v) b_T(v) \\ & - B_T^{-q} \int_{-T}^{T} dv e^{-ivw} (1 - k(B_T v)) R(v) \\ & - \frac{1}{T} B_T^{-q} \int_{-T}^{T} dv e^{-ivw} \mid v \mid k(B_T v) R(v) \\ & - B_T^{-q} \int_{\mid v \mid \geqq T} dv e^{-ivw} R(v). \end{aligned} \tag{5.13}$$

We now show that the first, third, and fourth terms in (5.13) tend to 0, as $T \to \infty$, uniformly in w. From (5.9) and (5.11), it follows that, uniformly in w,

$$\lim_{T\to\infty} E \left| \frac{1}{B_T^q} \int_{-T}^{T} dv e^{-ivw} k(B_T v) b_T(v) \right|^2 = 0.$$

Next, if M is an upper bound for $\mid k(v) \mid$, the third term in (5.13) tends to 0 by (5.7) and (5.9) if $q \geqq 1$, and if $q < 1$ it is in absolute value less than

$$\frac{M}{(TB_T)^q} \frac{1}{T^{1-q}} \int_{-T}^{T} dv \mid v \mid \mid R(v) \mid,$$

which tends to 0 by (5.7) and Lemma 4. Similarly, the fourth term in (5.13), which is in absolute value less than

$$\frac{1}{(TB_T)^q}\, T^q \int_{|v| \geqq T} dv \mid R(v) \mid,$$

tends to 0.

Consequently, (5.10) is proved for the case that the kernel has characteristic exponent $r = q$, since the second term in (5.13) then tends, uniformly in w, to $-2\pi\, k^{(q)} f^{(q)}(w)$. Next, to prove (5.10) for the case that $r > q$, it suffices to show that then

$$\lim_{T \to \infty} B_T^{-q} \int_{-T}^{T} \mid 1 - k(B_T v) \mid \mid R(v) \mid dv = 0. \tag{5.14}$$

Let M, M_1, and D be positive constants such that $\mid k(v) \mid \leqq M$ for all v, and

$$\mid 1 - k(v) \mid \leqq M_1 \mid v \mid^r \qquad \text{for } \mid v \mid \leqq D. \tag{5.15}$$

If the characteristic exponent is infinite, we may take any exponent $r > q$ in (5.15). Let $s = r - q$. Then the quantity in (5.14) whose limit is being taken is less than

$$M_1 B_T^s \int_{|v| \leqq DB_T^{-1}} \mid v \mid^{q+s} R(v)\, dv + M B_T^{-q} \int_{|v| \geqq DB_T^{-1}} R(v)\, dv,$$

which tends to 0 in view of (5.7) and Lemma 4.

We next establish Theorem 5A. In view of the foregoing, it follows that the desired asymptotic covariance in (5.6) is given by the limit, as $T \to \infty$, of

$$\frac{4}{4\pi^2}\, TB_T \int_0^T \int_0^T dv_1 dv_2 \cos w_1 v_1 \cos w_2 v_2 k(B_T v_1) k(B_T v_2)\; ED_T(v_1) D_T(v_2). \tag{5.16}$$

We may write (5.16) as a sum of three 3-fold integrals, by replacing $\mathrm{TED}_T(v_1)$ $D_T(v_2)$ by its value (3.10). The term in this sum which involves $Q(v_1, u, u + v_2)$ clearly vanishes in the limit, uniformly in w_1 and w_2.

Next we show that the term involving $R(u + v_1)\, R(u - v_2)$ also vanishes in the limit, uniformly in w_1 and w_2. For this term is less than

$$B_T \int_0^T dv_2 \int_0^T dv_1 \int_{-T}^{T} du \mid k(B_T v_1) k(B_T v_2) R(u + v_1) R(u - v_2) \mid$$

Making the change of variable $v_1 = z_1 - v_2$, $u = z + v_2$, this becomes

$$B_T \int_0^T dv_2 \int_{v_2}^{T+v_2} dz_1 \int_{-T-v_2}^{T-v_2} dz \mid k(B_T v_2) k(B_T v_2 - B_T z_1) R(z) R(z + z_1) \mid.$$

Making the change of variable $z_2 = B_T v_2$, this becomes

$$\int_0^{B_T T} dz_2 \int_{(z_2/B_T)}^{T+(z_2/B_T)} dz_1 \int_{-T[1+(z_2/TB_T)]}^{T[1-(z_2/TB_T)]} dz \mid k(z_2) k(z_2 - B_T z_1) R(z) R(z + z_1) \mid,$$

which tends to 0 as $T \to \infty$, since the region of integration over the z_1 variable tends to infinity.

The value of (5.6) is then given by the limit of

$$\frac{1}{\pi^2} B_T \int_0^T \int_0^T dv_1 dv_2 \cos w_1 v_1 \cos w_2 v_2 k(B_T v_1) k(B_T v_2) \int_{-T}^{T} du U_T(u, v_1, v_2) R(u) R(u + v_1 - v_2). \tag{5.17}$$

By the change of variables $u_1 = v_1 - v_2$, $u_2 = v_2$ this becomes

$$\frac{1}{\pi^2} B_T \int_0^T du_2 \int_{-u_2}^{T-u_2} du_1 \cos w_1(u_1 + u_2) \cos w_2 u_2 \, k(B_T u_2) k(B_T u_2 + B_T u_1) \int_{-T}^{T} du U_T(u, u_1 + u_2, u_2) R(u) R(u + u_1). \tag{5.18}$$

By the change of variable $z = B_T u_2$, and the formula $2 \cos A \cos B = \cos (A + B) + \cos (A - B)$, one obtains that (5.18) is equal to

$$\frac{1}{2\pi^2} \int_0^{B_T T} dz \int_{-(z/B_T)}^{T[1-(z/TB_T)]} du_1 \left\{ \cos \left[z \left(\frac{w_1 - w_2}{B_T} \right) + u_1 w_1 \right] + \cos \left[z \left(\frac{w_1 + w_2}{B_T} \right) + u_1 w_1 \right] \right\} k(z) k(z + B_T u_1) \int_{-T}^{T} du U_T \left(u, u_1 + \frac{z}{B_T}, \frac{z}{B_T} \right) R(u) R(u + u_1). \tag{5.19}$$

By referring to (3.11), it may be verified that, as

$$T \to \infty, \qquad U_T \left(u, u_1 + \frac{z}{B_T}, \frac{z}{B_T} \right) \to 1.$$

Now to evaluate (5.19), one may distinguish three cases: case I, $w_1 \neq w_2$; case II, $w_1 = w_2 = w \neq 0$; case III, $w_1 = w_2 = 0$. In view of the Riemann-Lebesque Lemma, the first term in (5.19) vanishes in the limit if $w_1 - w_2 \neq 0$, and the second term vanishes in the limit if $w_1 + w_2 \neq 0$. Further, for any $\epsilon > 0$, the convergence to 0 is uniform in w_1 and w_2 such that $w_1 \geqq \epsilon$ and $w_2 \geqq \epsilon$. Thus one obtains that, in the limit, the value of (5.19) is 0 in case I; in case II, it is equal to

$$\frac{1}{2\pi^2} \int_0^\infty k^2(z) \, dz \int_{-\infty}^{\infty} du_1 \cos w u_1 \int_{-\infty}^{\infty} du \, R(u) R(u + u_1); \tag{5.20}$$

and, in case III, it is equal to twice (5.20). It is easily verified that (5.20) and (5.6) are equal.

To adapt the foregoing argument to the discrete parameter case requires some care in the phase of the argument following (5.19). The integration in (5.19) involving the variable z should be replaced by a summation over the lattice points $z_j = jB_T$, where $j = 1, \cdots, T$. As $T \to \infty$, the distance between the lattice points tends to 0, and the highest lattice point tends to infinity, so

that the summation may be regarded as approaching the integral $\int_0^\infty k^2(z)\,dz$, as above.

6. Optimum consistent estimates of the spectral density. In view of Theorems 5A and 5B, the means are now at hand for choosing that estimate $f_T^*(w)$, of the form of (5.4), which is optimum in the sense that it possesses an order of consistency not less than that of any other such estimate. We obtain the following theorem.

THEOREM 6: *Suppose that* (5.7) *holds. Let the constants* B_T *be chosen so that, for some finite positive number* B,

$$\lim_{T\to\infty} T^{(1/1+2q)} B_T = B. \tag{6.1}$$

Let

$$\alpha = \frac{q}{1+2q}. \tag{6.2}$$

Then for any covariance averaging kernel $k(v)$ *with characteristic exponent* $r \geqq q$, *the corresponding estimate* $f_T^*(w)$ *possesses an asymptotic mean square error given by*

$$\lim_{T\to\infty} T^{2\alpha} E \mid f_T^*(w) - f(w) \mid^2 = \frac{f^2(w)}{B} \int k^2(z)\,dz\{1 + \delta(0, w)\} + B^{2q} \mid k^{(q)} f^{(q)}(w) \mid^2. \tag{6.3}$$

REMARK. If $q < r$, then $k^{(q)} = 0$.

Now, as q increases, the exponent α, as defined by (6.2), increases from 0 to $\frac{1}{2}$. Thus the factor which determines the highest order of consistency which may be attained, is the largest positive number q such that (5.7) holds. For want of a better name, we call this largest q the *exponent of uncorrelation* of the time series whose covariance function is $R(v)$, since the larger q is, the faster $R(v)$ falls off as $v \to \infty$, and the less correlated are successive observations of the time series. If (5.7) holds for all finite values of q, as is the case if $R(v)$ decreases exponentially, we define the exponent of uncorrelation to be infinite.

For computational convenience, the kernel with characteristic exponent r that we prefer is

$$\begin{aligned} k_r(z) &= 1 - \mid z \mid^r \qquad \text{if } \mid z \mid < 1, \\ &= 0 \qquad\qquad\quad \text{otherwise.} \end{aligned} \tag{6.4}$$

Such a kernel leads to an estimate which does not require the computation of all the sample covariances. With this choice of kernel, $f_T^*(w)$ may be written letting $M_T = (1/B_T) \leqq T$,

$$f_T^*(w) = \frac{1}{2\pi} \sum_{\mid v \mid \leqq M_T} e^{-iwv} \left\{1 - \left(\frac{\mid v \mid}{M_T}\right)^r\right\} R_T(v). \tag{6.5}$$

The foregoing results may be interpreted from two points of view, emphasizing either the choice of kernel $k(z)$ or the choice of constants $M_T = 1/B_T$ (which,

in the case of a kernel vanishing for $|z| > 1$, represent the number of sample covariances included in the estimate).

Let a kernel $k(z)$ be chosen whose characteristic exponent is r. Then the order of consistency of the corresponding estimate cannot be greater than $T^{2\alpha(r)}$, where $\alpha(r) = r/(1 + 2r)$, and will be $T^{2\alpha}$, where $\alpha \leqq \alpha(r)$, if the constants M_T satisfy the relation for some finite positive number M,

$$\lim_{T\to\infty} \frac{M_T}{T^{1-2\alpha}} = M \tag{6.6}$$

and if (5.7) holds for $q = \alpha/(1 - 2\alpha)$.

Therefore, if Bartlett's modified periodogram (which is (6.5) with $r = 1$) is used as the estimate, its order of consistency cannot be greater than $T^{2/3}$, and will be $T^{2\alpha}$ (where $\alpha \leqq \frac{1}{3}$) if the number of sample covariances included in the estimate is $T^{1-2\alpha}$. If the truncated periodogram (which is (6.5) with $r = \infty$) is used as the estimate, its order of consistency will be $T^{2\alpha}$ (where $\alpha < \frac{1}{2}$) if $M_T = T^{1-2\alpha}$, *and* if (5.7) holds for $q = \alpha/(1 - 2\alpha)$, which would be the case if the exponent of uncorrelation is infinite.

On the other hand, let the constants M_T be chosen so that (6.6) holds for some α between 0 and $\frac{1}{2}$. Then the order of consistency of the corresponding estimate $f_T^*(w)$ is $T^{2\alpha}$, no matter what the value of the characteristic exponent r of the kernel used so long as $r \geqq q(\alpha) = \alpha/(1 - 2\alpha)$, and (5.7) holds for $q = q(\alpha)$.

7. Alternative estimates of the spectral averages. In our study of the consistent estimates of the spectral density, we were led to consider estimates, such as Bartlett's modified periodogram, which had the property of only requiring the computation, on the basis of an observed sample of length T, of the sample covariances $R_T(v)$ for $|v|$ less than some root of T. In this section we show that for the spectral averages $J(A)$, one may define estimates $J_T^*(A)$, alternative to the previously given estimates $J_T(A)$, which have the same order of consistency and asymptotic variance as the latter, and require the computation of fewer sample covariances.

Let $A(w)$ be a spectral averaging function, with Fourier transform $a(v)$. Let $k(z)$ be a covariance averaging kernel, with Fourier transform $K(w)$. Let B_T be a sequence of constants tending to 0. Let $f_T^*(w)$ be defined by (5.4). Define

$$J_T^*(A) = \int f_T^*(w)A(w)\,dw. \tag{7.1}$$

One may write $J_T^*(A)$ in terms of the sample spectral density function by

$$J_T^*(A) = \int f_T(w)A_T(w)\,dw, \tag{7.2}$$

where

$$A_T(w) = \frac{1}{B_T}\int K\left(\frac{w - \lambda}{B_T}\right)A(\lambda)\,d\lambda.$$

In terms of the sample covariance functions, one may write

$$J_T^*(A) = \int_T^T a(v)k(B_T v)R_T(v)\,dv \tag{7.3c}$$

$$= \sum_{|v|\leq T} a(v)k(B_T v)R_T(v). \tag{7.3d}$$

The properties of the estimate $J_T^*(A)$ are embodied in the following two theorems, whose proofs are omitted.

THEOREM 7A. *For any two spectral averaging functions $A_1(w)$ and $A_2(w)$, the covariance* $\mathrm{Cov}[J_T^*(A_1), J_T^*(A_2)]$ *satisfies* (4.11).

THEOREM 7B. *Let $a(v)$ be a covariance averaging function. Let $q > \frac{1}{2}$ be such that*

$$\int |v|^q \, |a(v)R(v)|\,dv < \infty. \tag{7.4}$$

Let $k(z)$ be a covariance averaging kernel with characteristic exponent $r \geq q$. Let the positive constants B_T be chosen so that

$$\limsup_{T\to\infty} T^{1/2}B_T^q \begin{cases} = 0 & \text{if } r = q, \\ < \infty & \text{if } r > q. \end{cases} \tag{7.5}$$

Then the bias $E\,J_T^*(A) - J(A)$ satisfies (4.10).

Optimum Estimates: The estimates $J_T^*(A)$ are all equivalent from the point of view of their order of consistency and asymptotic variance. If one desires to choose between them, the only basis is computational convenience. It is with this in mind that the following remarks are made. For the covariance averaging kernel, we choose $k_r(z)$. Then (7.3d) becomes, letting $M_T = (1/B_T)$

$$J_T^*(A) = \sum_{|v|\leq M_T} a(v)\left\{1 - \left(\frac{|v|}{M_T}\right)^r\right\} R_T(v). \tag{7.7}$$

We choose B_T to be of the form $B_T = T^{-m}$, where the positive exponent m is to be chosen as small as possible, so that the number of terms in (7.7) will be as small as possible. Let q be the largest positive number such that (7.4) holds. Assuming q to be finite, choose $r \geq q$. Then $J_T^*(A)$ will give a consistent estimate of $J(A)$, involving the calculation of a minimum number of sample covariances, if m is chosen as near to the lower bound as possible in the inequalities

$$\begin{aligned} m &> \frac{1}{2q} \quad \text{if } r = q, \\ m &\geq \frac{1}{2q} \quad \text{if } r > q. \end{aligned} \tag{7.8}$$

8. Interpolating the spectral density. In order to obtain an estimate of the complete graph of the spectral density function $f(w)$ by means of the estimates $f_T^*(w)$ discussed in the foregoing, one needs to compute the estimate at all values of w. In this section, estimates $f_T^{**}(w)$ are constructed, which are equivalent to $f_T^*(w)$ from the point of view of order of consistency and asymptotic variance,

and which require the computation of only a finite number of quantities in order to obtain the entire graph. Only the discrete parameter case is discussed in detail.

To begin with, define

$$w_m(T) = \frac{2\pi m}{2T+1} \qquad \text{for } m = 0, \pm 1, \cdots, \pm T, \tag{8.1}$$
$$= 0 \qquad \text{otherwise.}$$

We now show that $f_T^*(w)$, as defined by (5.4), may be expressed in terms of its values at the above $(2T+1)$ lattice points by the formula

$$f_T^*(w) = \sum_{m=-T}^{T} c_m(w; T) f_T^*(w_m(T)), \tag{8.2}$$

where

$$c_m(w; T) = \frac{1}{2T+1} \sum_{v=-T}^{T} \exp\,[-iv(w - w_m(T))] \tag{8.3}$$
$$= \frac{\sin\,[(1/2)(2T+1)(w - w_m(T))]}{(2T+1)\sin\,[(1/2)(w - w_m(T))]}$$

To prove (8.2), we note that, for $v = 0, \pm 1, \cdots, \pm T$ and any w,

$$e^{-iwv} = \sum_{m=-T}^{T} c_m(w; T) \exp\,[-ivw_m(T)], \tag{8.4}$$

which may be verified by expanding the right-hand side. It is now easy to obtain (8.2) by substituting (8.4) into (5.4).

If $f_T^*(w)$ is given by (6.5), then it is determined by its value at even a fewer number of points, namely the lattice points $w_m(M_T)$, since by the same argument as above we may write

$$f_T^*(w) = \sum_{|m| \leq M_T} c_m(w; M_T) f_T^*(w_m(M_T)). \tag{8.5}$$

Thus it is seen that it suffices to compute $f_T^*(w)$ at a finite number of points in order to know it on the entire interval $0 \leqq w \leqq \pi$. In view of the peaked nature of the $c_m(w; T)$ functions for large T, it might be thought that an adequate approximation to $f_T^*(w)$ would be $f_T^*(w_m(T))$, where $w_m(T)$ is the lattice point nearest to w. The problem which is raised by the representations (8.2) and (8.5) is when is such an approximation valid. From a statistical point of view, the justification of such an approximation must be in terms of its providing an estimate which has the proper order of consistency and asymptotic variance. It is from this point of view that we now consider the problem of using the value of $f_T^*(w)$ at a finite number of points to obtain estimates of its value at all points.

Let d_T be a sequence of constants tending to 0, and define the function, for $w \geqq 0$,

$$W_T(w) = \left[\frac{w}{d_T}\right] d_T, \tag{8.6}$$

where $[x]$ denotes the largest integer smaller than x. Consider the following estimate of the spectral density function

$$f_T^{**}(w) = f_T^{*}(W_T(w)), \tag{8.7}$$

where $f_T^{*}(w)$ is defined by (5.4). This estimate clearly has only a finite number of distinct values. The properties of this estimate are embodied in the following theorem.

THEOREM 8: *Assume that the conditions of Theorem* 6 *are fulfilled, so that the estimate* $f_T^{*}(w)$ *is consistent of order* $T^{2\alpha}$, *with asymptotic variance given by* (6.3). *Let*

$$\beta = \frac{\alpha}{q} = \frac{1}{1+2q} = 1 - 2\alpha. \tag{8.8}$$

Let the positive constants d_T *be chosen such that*

$$\limsup_{T\to\infty} T^{\beta} d_T < \infty \qquad \text{if } 0 < q < 1, \qquad \text{whence } 0 < \alpha < \tfrac{1}{4}, \tag{8.9}$$

$$\lim_{T\to\infty} T^{\alpha} d_T = 0 \qquad \text{if } q \geqq 1, \qquad \text{whence } \tfrac{1}{4} \leqq \alpha < \tfrac{1}{2}. \tag{8.10}$$

Then the estimate $f_T^{**}(w)$ *is consistent of order* $T^{2\alpha}$, *with the same asymptotic bias and asymptotic variance as* $f_T^{*}(w)$.

PROOF. One may suppose $w > 0$, since $f_T^{**}(0) = f_T^{*}(0)$. Now $w - d_T \leqq W_T(w) \leqq w$, so that $W_T(w) \to w$. In view of the uniform convergence in (5.6), it follows that the asymptotic variance of $f_T^{**}(w)$ is the same as that of $f_T^{*}(w)$. Next, in view of the uniform convergence in (5.10), to establish that $f_T^{**}(w)$ has the same asymptotic bias as that of $f_T^{*}(w)$ it suffices to show that

$$\lim_{T\to\infty} T^{\alpha} \left| f(W_T(w)) - f(w) \right| = 0. \tag{8.11}$$

Now the quantity in (8.11) whose limit is being taken is less than

$$T^{\alpha} d_T \sum_{|v| \leq T^{\beta}} |v| \, R(v) + 2T^{\alpha} \sum_{|v| \geq T^{\beta}} |R(v)|.$$

The second term tends to 0, since it is less than

$$2T^{\alpha-q} \sum_{|v| \geq T^{\beta}} |v|^{q} |R(v)|.$$

The first term also tends to 0; by (8.10) if $q \geqq 1$, and by Lemma 4 and (8.9), if $q < 1$, since the term may be written

$$T^{\alpha-\beta q} T^{\beta} d_T \left(\frac{1}{T^{\beta}}\right)^{1-q} \sum_{|v| \leq T^{\beta}} |v| \, |R(v)|.$$

If d_T is chosen by

$$d_T = \pi T^{-1/2}, \tag{8.12}$$

then (8.9) and (8.10) will be satisfied for $\frac{1}{4} \leqq \alpha < \frac{1}{2}$ (which corresponds to $\frac{1}{2} \leqq q < \infty$). It would seem that (8.12) provides a safe universal choice of the spacing of the lattice points.

If it is desired that the estimate of the spectral density be a continuous function, without jumps, then one may use the estimate

$$f_T^{**}(w) = a_T f_T^*(W_T(w)) + (1 - a_T) f_T^*(W_T(w) + d_T), \tag{8.13}$$

where a_T is a sequence of constants between 0 and 1, not in general approaching a limit. It may be verified that Theorem 8 holds for the estimate given by (8.13), provided that in (8.9) it is required that the limit be 0. Then it follows that $(d_T/B_T) \to 0$, and

$$\lim_{T\to\infty} T^{2\alpha} \operatorname{Cov}[f_T^*(W_T(w)), f_T^*(W_T(w) + d_T)] = \lim_{T\to\infty} T^{2\alpha} \operatorname{Var}[f_T^*(w)].$$

REFERENCES

[1] M. S. Bartlett, "Periodogram analysis and continuous spectra," *Biometrika*, Vol. 37 (1950), pp. 1–16.

[2] M. S. Bartlett, *An Introduction to Stochastic Processes*, Cambridge, 1955.

[3] U. Grenander, "On empirical spectral analysis of stochastic processes," *Ark. Mat.*, Vol. 1 (1951), pp. 503–531.

[4] U. Grenander and M. Rosenblatt, "Statistical spectral analysis of time series arising from stationary stochastic processes," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 537–559.

[5] U. Grenander and M. Rosenblatt, "Regression analysis of time series with stationary residuals," *Proc. Nat. Acad. Sci.*, Vol. 40 (1954), pp. 812–816.

[6] T. A. Magness, "Spectral response of a quadratic device to non-Gaussian noise," *J. Appl. Phys.*, Vol. 25 (1954), pp. 1357–1365.

[7] J. Tukey, "Measuring Noise Color," an unpublished manuscript prepared for distribution at a meeting of the Institute of Radio Engineers, November 7, 1951.

[8] E. Parzen, *Optimum Methods of Spectral Analysis of Finite Noise Samples*, Columbia University, Hudson Laboratories Technical Report No. 39, issued Summer 1956.

[9] J. L. Doob, *Stochastic Processes*, New York, 1953.

[10] M. Loève, *Probability Theory*, New York, 1955.

INTRA-BLOCK ANALYSIS FOR FACTORIALS IN TWO-ASSOCIATE CLASS GROUP DIVISIBLE DESIGNS[1, 2]

CLYDE YOUNG KRAMER AND RALPH ALLAN BRADLEY

Virginia Polytechnic Institute

1. Introduction and summary. Group divisible incomplete block designs form an important class of incomplete block designs useful in a wide variety of experimental situations. Their properties, construction, and analysis have been thoroughly discussed in statistical literature, and we cite only several recent references [1], [2], and [3] to work of Bose and his co-workers dealing with partially balanced designs with two associate classes with which we shall be concerned.

The utility of incomplete block designs would be increased with means of incorporating factorial treatment combinations in them. The use of factorials is widespread and stimulated by the concepts of confounding, partial confounding, and fractional replication. A mathematical summary on factorials is given by Kempthorne [4]. Kramer and Bradley [5] considered factorials in near-balance incomplete block designs, and here we generalize to the wider class of group divisible designs with two associate classes. Harshbarger [6] used a 2^3 factorial in a Latinized rectangular lattice and this seems to be the first use of a factorial in a partially balanced incomplete block design.

We obtain the intra-block analysis of variance for two-associate class group divisible designs with the adjusted treatment sum of squares in a modified form that more clearly indicates the structure of that quantity. Factorial treatment combinations are then identified with basic treatments through the association scheme of a design. This identification is effected in such a way that the factors are divided into two groups. For example, the design for 18 treatments (see [2], Design S60), divisible into six groups of three, in blocks of six, treatments replicated five times, can be adapted to a 6×3 factorial scheme; by regarding the six groups as made up of a 2×3 classification, the same design can be used for a 2×3^2 factorial scheme. Single-degree-of-freedom comparisons are obtainable in much the usual way and use of fractional replication, essentially within the groups of factors, is possible. The analyses for factorials depend on the estimators of basic treatment effects.

We are not concerned with the construction of two-associate class group divisible designs and all known such designs for which $r \leqq 10$, $3 \leqq k \leqq 10$, where r is the number of replications and k is the number of plots per block, are given in [7].

Received July 23, 1956; revised October 12, 1956.

[1] Research sponsored by the Agricultural Research Service, U.S.D.A., under a Research and Marketing Act Contract, No. 12-14-100-126(20).

[2] A condensation of part of a dissertation by C. Y. Kramer, written under the direction of R. A. Bradley and submitted to the Virginia Polytechnic Institute in partial fulfillment of the requirements for the Ph.D. degree in Statistics.

2. Definitions and notation. Bose, Clatworthy, and Shrikhande [7] list the following properties of group divisible designs with two associate classes:

(i) The experimental material is divided into b blocks of k units each; different treatments are applied to the different units in a block.

(ii) There are $v = mn$ treatments ($v > k$) and the treatments can be divided into m groups of n each such that any two treatments of the same group are first associates while two treatments from different groups are second associates. Each treatment occurs in the design r times, and $vr = bk$.

(iii) Each treatment has exactly $(n - 1)$ first associates and $n(m - 1)$ second associates.

(iv) Given any two treatments which are ith associates, the number of treatments common to the jth associate of the first and the kth associate of the second is p_{jk}^{i}, $(i, j, k = 1, 2)$, and is independent of the pair of treatments selected. In matrix notation, if P_i is the matrix with elements p_{jk}^{i},

$$P_1 = \begin{Vmatrix} (n-2) & 0 \\ 0 & n(m-1) \end{Vmatrix} \text{ and } P_2 = \begin{Vmatrix} 0 & (n-1) \\ (n-1) & n(m-2) \end{Vmatrix}.$$

(v) Two treatments which are ith associates occur together in exactly λ_i blocks, $i = 1, 2$.

(vi) The inequalities, $r \geqq \lambda_1$, $rk - \lambda_2 v \geqq 0$, hold.

(vii) The design parameters are related so that $(n - 1)\lambda_1 + n(m - 1)\lambda_2 = r(k - 1)$, or $rk - \lambda_2 v = r - \lambda_1 + n(\lambda_1 - \lambda_2)$. Group divisible designs have been divided into three subclasses, Singular, Semi-regular, and Regular, but we shall consider the class as a whole without subdivision.

We let V_{ij} denote the jth treatment of the ith group noted in (ii), $i = 1, \cdots, m$; $j = 1, \cdots, n$. Then the usual association scheme is given by the matrix V with elements V_{ij}. Two treatments with common first subscripts (in the same row of V) are first associates; otherwise they are second associates. The double subscript notation is introduced here for it will be convenient when we come to consider factorials. To use the design catalogue [7] it is only necessary to match our treatment designations with those in the association matrices where treatments are numbered serially.

The model that will be assumed for group divisible incomplete block designs is that

$$y_{ijs} = \mu + \tau_{ij} + \beta_s + \epsilon_{ijs}, \tag{2.1}$$

where y_{ijs} is the observation on V_{ij} in block s if that treatment occurs in block s, μ is the grand mean, τ_{ij} is the effect of V_{ij}, β_s is the effect of block s, and ϵ_{ijs} are independent normal variates with zero means and homogeneous variances, σ^2. Latin letters m, t_{ij} and b_s will be used for estimators of the parameters in (2.1). Restrictions on the parameters in (2.1) are

$$\sum_i \sum_j \tau_{ij} = 0 \tag{2.2}$$

and

$$\sum_{\cdot} \beta_{\cdot} = 0. \tag{2.3}$$

The parameter $\beta_{\cdot}$ in (2.1) may sometimes be redefined when the blocks are arranged in replications or a Latin square [8], [9]. We shall not explicitly consider these situations since the modifications involved do not affect the estimation of the adjusted treatment sum of squares.

3. General regression theory. Let

$$y_\alpha = \mu + \sum_{i=1}^{k} \beta_i x_{i\alpha} + \epsilon_\alpha, \tag{3.1}$$

$\alpha = 1, \cdots, N$, represent a general regression model where the $x_{i\alpha}$ are constants and the ϵ_α are independent normal variates with zero means and homogeneous variances, σ^2. The β_i are regression parameters subject to r_1 linearly independent restrictions,

$$\sum_{i=1}^{k} \alpha_{ih}\beta_i = 0, \qquad h = 1, \cdots, r_1 < (k-1), \tag{3.2}$$

defining a parameter space Ω. The α_{ih} in (3.2) are known constants. A null hypothesis introduces r_2 additional restraints through additional linearly independent equations like (3.2) for $h = r_1 + 1, \cdots, r_1 + r_2 < (k-1)$ and thus defines a parameter space ω, a subspace of Ω.

The general theory of regression tests under the conditions set forth lets us state that $\text{Reg}\,(\beta \mid \Omega)/\sigma^2$, $\text{Reg}\,(\beta \mid \omega)/\sigma^2$, and $[\text{Reg}\,(\beta \mid \Omega) - \text{Reg}\,(\beta \mid \omega)]/\sigma^2$ have χ^2-distributions respectively with $(k - r_1)$, $(k - r_1 - r_2)$, and r_2 degrees of freedom independent of $\text{Res}\,(\beta \mid \Omega)/\sigma^2$, which also has a χ^2-distribution with $(N - k + r_1 - 1)$ degrees of freedom. $\text{Reg}\,(\beta \mid \Omega)$ is the sum of squares due to regression on the x-variables in (3.1) with the regression coefficients subject to the restraints (3.2); $\text{Res}\,(\beta \mid \Omega)$ is the resultant sum of squares of deviations about that regression line. $\text{Reg}\,(\beta \mid \omega)$ is the sum of squares due to regression on the x-variables in (3.1) with the regression coefficients subject to the totality of $(r_1 + r_2)$ restraints defining ω. We note that

$$\text{Reg}\,(\beta \mid \Omega) = \sum_{i=1}^{k} b_i g_i, \tag{3.3}$$

$$\text{Reg}\,(\beta \mid \omega) = \sum_{i=1}^{k} b_i' g_i, \tag{3.4}$$

and

$$\text{Res}\,(\beta \mid \Omega) = \sum_{\alpha=1}^{N} (y_\alpha - \bar{y})^2 - \sum_{i=1}^{k} b_i g_i, \tag{3.5}$$

where

$$g_i = \sum_{\alpha=1}^{N} (y_\alpha - \bar{y}) x_{i\alpha}, \tag{3.6}$$

b_i and b_i' are the least squares estimators of β_i under the restraints of Ω and ω respectively, and $\bar{y} = \sum_{\alpha=1}^{N} y_\alpha/N$. An F-test of the indicated hypothesis is possible based on

$$F = (N - k + r_1 - 1)[\text{Reg } (\beta \mid \Omega) - \text{Reg } (\beta \mid \omega)]/r_2 \text{ Res } (\beta \mid \Omega), \tag{3.7}$$

with r_2 and $(N - k + r_1 - 1)$ degrees of freedom.

To illustrate the application of this theory, we consider the model (2.1) corresponding to (3.1) and the restrictions (2.2) and (2.3) corresponding to (3.2) and defining Ω. Now N in the regression theory is replaced by vr, k by $(b + v)$, r_1 by 2, and r_2 by $(v - 1)$. The regression coefficients β_i become treatment and block effects, τ_{ij} and β_s. To test the hypothesis that $\tau_{ij} = 0$ for all i and j in (2.1), the hypothesis of "no treatment effects", it is only necessary to add $(v - 1)$ additional linearly independent restrictions on the τ_{ij} to insure that each $\tau_{ij} = 0$, thus defining ω. The adjusted treatment sum of squares with $(v - 1)$ degrees of freedom becomes

$$\text{Reg } (\beta, \tau \mid \Omega) - \text{Reg } (\beta, \tau \mid \omega), \tag{3.8}$$

where

$$\text{Reg } (\beta, \tau \mid \Omega) = \sum_i \sum_j t_{ij} T_{ij} + \sum_s b_s B_s \tag{3.9}$$

and

$$\text{Reg } (\beta, \tau \mid \omega) = \sum_s b_s' B_s\,, \tag{3.10}$$

the latter sums of squares having respectively $(b + v - 2)$ and $(b - 1)$ degrees of freedom. T_{ij} is the total for treatment V_{ij} and B_s is the sth block total. b_s and b_s' are the estimators of β_s under Ω and ω respectively; t_{ij}', the estimator of τ_{ij} under ω, is necessarily zero. The error sum of squares for the intra-block analysis of variance is

$$\text{Res } (\beta, \tau \mid \Omega) = \sum_i \sum_j \sum_s (y_{ijs} - \bar{y})^2 - \text{Reg } (\beta, \tau \mid \Omega), \tag{3.11}$$

with $(vr - b - v + 1)$ degrees of freedom. In (3.11), note that the summation is restricted to values of i and j occurring with s through the properties of the designs considered; this will be the case throughout this paper. The unadjusted block sum of squares is given by (3.10) and has $(b - 1)$ degrees of freedom.

We shall use the theory summarized in this section in the subsequent discussions. A basis for this theory is given by Wilks ([10], Sections 8.3 and 8.43).

4. General analysis of variance modified. The basic intra-block analysis of variance for partially balanced incomplete block designs with two-associate classes is known ([7], Table 1.0). In our notation, the adjusted treatment sum of squares is

$$\text{Adj. Treat. S.S.} = \sum_i \sum_j t_{ij} Q_{ij}\,, \tag{4.1}$$

where

$$Q_{ij} = T_{ij} - B_{ij\cdot}/k, \tag{4.2}$$

with $B_{ij.}$, the total of block totals for blocks containing V_{ij}. For the subclass of group divisible designs,

$$(4.3) \qquad v\lambda_2(\lambda_1 + rk - r)t_{ij} = k(\lambda_1 + \lambda_2 v - \lambda_2)Q_{ij} + k(\lambda_1 - \lambda_2)\sum_{\substack{p \\ p \neq j}} Q_{ip}$$

obtained from the reference ([7], Eqs. 1.11 to 1.19). If j in (4.3) is replaced by q and both sides of (4.3) summed over values of $q \neq j$, the resulting identity may be substituted back into (4.3) with simple algebraic reduction based on the relations (vii) of Section 2 to yield

$$(4.4) \qquad [(\lambda_2 + rk - r)t_{ij} + (\lambda_2 - \lambda_1)\sum_{\substack{p \\ p \neq j}} t_{ip}]/k = Q_{ij}.$$

The adjusted treatment sum of squares expressed in terms of the estimators of treatment effects alone is

$$(4.5) \qquad \text{Adj. Treat. S.S.} = \frac{(\lambda_1 + rk - r)}{k}\sum_i \sum_j t_{ij}^2 + \frac{(\lambda_2 - \lambda_1)}{k}\sum_i \left(\sum_j t_{ij}\right)^2,$$

obtained by substituting Q_{ij} in (4.4) into (4.1).

The result of (4.5) is a form more suitable for the consideration of factorials than (4.1). Usually in analysis of variance, computing is based on (4.1). It is in fact simpler when using a desk calculator to substitute for the Q_{ij} in (4.3) to obtain

$$(4.6) \qquad \begin{aligned} t_{ij} = [kv\lambda_2 T_{ij} - k(\lambda_2 - \lambda_1)\textstyle\sum_j T_{ij} - v\lambda_2 B_{ij.} \\ + (\lambda_2 - \lambda_1)\textstyle\sum_j B_{ij.}]/v\lambda_2(\lambda_1 + rk - r) \end{aligned}$$

from two-way tables of values of T_{ij} and $B_{ij.}$. Substitution in (4.5) is then based on (4.6).

The analysis of variance is completed by the calculation of the unadjusted block sum of squares and the total sum of squares, for the error sum of squares is obtained by subtraction.

$$(4.7) \qquad \text{Unadj. Block S.S.} = \frac{1}{k}\sum_s B_s^2 - \frac{G^2}{rv}.$$

$$(4.8) \qquad \text{Total S.S.} = \sum_i \sum_j \sum_s y_{ijs}^2 - \frac{G^2}{rv}.$$

G is the grand total of all observations, $\sum_i \sum_j \sum_s y_{ijs}$. Degrees of freedom for Adj. Treat. S.S., Unadj. Block S.S., Total S.S., and Error S.S. are respectively $(v - 1)$, $(b - 1)$, $(rv - 1)$, and $[(r - 1)v - b + 1]$.

The variance of the difference between estimators of first-associate treatment effects is

$$(4.9) \qquad V(t_{ij} - t_{ij'}) = 2k\sigma^2/(\lambda_1 + rk - r),$$

$j \neq j'$; the variance of the difference between estimators of second-associate treatment effects is

$$V(t_{ij} - t_{i'j'}) = 2k\sigma^2(\lambda_1 + \lambda_2 v - \lambda_2)/v\lambda_2(\lambda_1 + rk - r), \tag{4.10}$$

$i \neq i'$. These variances are estimated by substituting the error mean square from the analysis of variance for σ^2.

The efficiencies of first and second associate treatment comparisons have been given by Bose and his associates [7]. These efficiencies are obtained by taking the ratio of the variance of the treatment contrast for a randomized block design to the corresponding variance for the incomplete block design given equal values of r and on the assumption that both designs yield the same experimental error. The efficiency for the comparison of two treatments that are first associates is

$$E_1 = (\lambda_1 + rk - r)/rk \tag{4.11}$$

and, for two treatments that are second associates, the efficiency is

$$E_2 = v\lambda_2(\lambda_1 + rk - r)/rk(\lambda_1 + \lambda_2 v - \lambda_2). \tag{4.12}$$

E_1 and E_2 are in more explicit forms than given previously and are derivable from (4.9) and (4.10) and the fact that the corresponding variance for the randomized block design is $2\sigma^2/r$.

5. The basic two-factor factorial. To introduce factorials into two-associate class group divisible designs, we first consider a basic two-factor factorial. Then it will be possible to show how multi-factor factorials may be used.

Consider A and C factors with m and n levels respectively providing $v = mn$ treatment combinations associated with the V_{ij} so that

$$\tau_{ij} = \alpha_i + \gamma_j + \delta_{ij} \tag{5.1}$$

with restrictions,

$$\textstyle\sum_i \alpha_i = 0, \tag{5.2}$$

$$\textstyle\sum_j \gamma_j = 0, \tag{5.3}$$

$$\textstyle\sum_i \delta_{ij} = 0, \tag{5.4}$$

and

$$\textstyle\sum_i \delta_{ij} = 0. \tag{5.5}$$

Equations (5.2) to (5.5) represent $(m + n + 1)$ linearly independent restrictions on the $(mn + m + n)$ new parameters. α_i, γ_j, and δ_{ij} are parameters representing respectively the effects of the ith level of the A-factor, the jth level of the C-factor and the interaction of the ith level of the A-factor and the jth level of the C-factor. Corresponding Latin letters will be used for estimators of these effects.

The change to factorial parameters may be regarded simply as a one-to-one transformation in the parameter space. It follows that

$$t_{ij} = a_i + c_j + d_{ij} \tag{5.6}$$

and substitution in (4.5) yields

$$\text{Adj. Treat. S.S.} = \frac{n\lambda_2 v}{k}\sum_i a_i^2 + \frac{m(\lambda_1 + rk - r)}{k}\sum_j c_j^2 + \frac{(\lambda_1 + rk - r)}{k}\sum_i\sum_j d_{ij}^2 \tag{5.7}$$

after reduction based on (vii) of Section 2. Use of the general regression theory is sufficient to obtain

$$\text{Adj. S.S. } (A) = \frac{n\lambda_2 v}{k}\sum_i a_i^2, \tag{5.8}$$

$$\text{Adj. S.S. } (C) = \frac{m(\lambda_1 + rk - r)}{k}\sum_j c_j^2, \tag{5.9}$$

and

$$\text{Adj. S.S. } (AC) = \frac{(\lambda_1 + rk - r)}{k}\sum_i\sum_j d_{ij}^2, \tag{5.10}$$

with $(m - 1)$, $(n - 1)$, and $(m - 1)(n - 1)$ degrees of freedom respectively. The complete analysis of variance is given in Table 1. Definition of (5.8), (5.9), and (5.10) is complete when we note that

$$a_i = \textstyle\sum_j t_{ij}/n = \bar{t}_{i\cdot}\,, \tag{5.11}$$

$$c_j = \textstyle\sum_i t_{ij}/m = \bar{t}_{\cdot j}\,, \tag{5.12}$$

and

$$d_{ij} = t_{ij} - \bar{t}_{i\cdot} - \bar{t}_{\cdot j}\,, \tag{5.13}$$

computed most easily from the two-way table of values of t_{ij}. Independence of the sums of squares in (5.8), (5.9), and (5.10) follows from Cochran's theorem [11].

We sketch the use of the general regression theory of Section 3 and the application of it to our problem by considering Adj. S.S. (A).

To effect the regression with the complete model obtained by substituting for τ_{ij} of (5.1) in (2.1), it is necessary to minimize

$$\textstyle\sum_i\sum_j\sum_s (y_{ijs} - \mu - \alpha_i - \gamma_j - \delta_{ij} - \beta_s)^2, \tag{5.14}$$

subject to the restraint (2.3) and to the $(m + n + 1)$ linearly independent restraints of (5.2) to (5.5) through use of Lagrange multipliers. The resulting

TABLE 1
Intra-block analysis of variance for the basic two-factor factorial

Source of Variation	Degrees of Freedom	Sum of Squares*
Treatments (adjusted)	$(v-1) = (mn-1)$	$K_1 \sum_i \sum_j t_{ij}^2 + K_2 \sum_i (\sum_j t_{ij})^2$
A-factor (adjusted)	$(m-1)$	$(nK_1 + n^2K_2) \sum_i \bar{t}_{i.}^2$
C-factor (adjusted)	$(n-1)$	$mK_1 \sum_j \bar{t}_{.j}^2$
AC-interaction (adjusted)	$(m-1)(n-1)$	$K_1 \sum_i \sum_j (t_{ij} - \bar{t}_{i.} - \bar{t}_{.j})^2$
Blocks (unadjusted)	$(b-1)$	$\frac{1}{k} \sum_s B_s^2 - \frac{G^2}{rv}$
Error	$[mn(r-1) - b + 1]$	By subtraction
Total	$(mnr-1)$	$\sum_i \sum_j \sum_s y_{ijs}^2 - \frac{G^2}{rv}$

* $K_1 = (\lambda_1 + rk - r)/k$, $K_2 = (\lambda_2 - \lambda_1)/k$, and $nK_1 + n^2K_2 = n\lambda_2 v/k$.

estimators are those given in (5.11) to (5.13) for α_i, γ_j, and δ_{ij}, and the estimator of μ is G/vr. It follows that

$$\text{Reg}\,(\alpha, \gamma, \delta, \beta \mid \Omega) = \sum_i a_i A_i + \sum_j c_j C_j + \sum_i \sum_j d_{ij} D_{ij} + \sum_s b_s B_s\,, \tag{5.15}$$

where $A_i = \sum_j T_{ij}$, $C_j = \sum_i T_{ij}$, $D_{ij} = T_{ij}$, and B_s is defined after (3.9). Ω is the parameter space defined by the indicated restrictions.

The null hypothesis of no A-effects implies $(m-1)$ additional linearly independent restrictions sufficient to make each $\alpha_i = 0$, and they reduce consideration to a parameter space ω_A, a subspace of Ω. Under these conditions it is necessary to minimize

$$\sum_i \sum_j \sum_s (y_{ijs} - \mu - \gamma_j - \delta_{ij} - \beta_s)^2 \tag{5.16}$$

with use of Lagrange multipliers and the restraints (2.3) and (5.3) to (5.5). Estimators of μ, γ_j, and δ_{ij} are unchanged; a new estimator b_s' of β_s is obtained. Now

$$\text{Reg}\,(\gamma, \delta, \beta \mid \omega_A) = \sum_j c_j C_j + \sum_i \sum_j d_{ij} D_{ij} + \sum_s b_s' B_s\,. \tag{5.17}$$

The estimators b_s and b_s' are fairly complex, but we need only note that

$$b_s' = b_s + \frac{1}{k} \sum_{\substack{i \\ \text{in } s}} n_s(i) a_i\,, \tag{5.18}$$

where $n_s(i)$ is the number of times a treatment combination with the ith level of the A-factor occurs in block s.

Adj. S.S. (A) is the difference, Reg $(\alpha, \gamma, \delta, \beta \mid \Omega)$ − Reg $(\gamma, \delta, \beta \mid \omega_A)$; and, using (5.15), (5.17), and (5.18), we have

$$\text{Adj. S.S. } (A) = \sum_i a_i A_i - \frac{1}{k} \sum_s \sum_{\substack{i \\ \text{in } s}} n_s(i) a_i B_s . \tag{5.19}$$

But

$$\sum_s \sum_{\substack{i \\ \text{in } s}} n_s(i) a_i B_s = \sum_i a_i \sum_{\substack{s \\ \text{with } i}} n_s(i) B_s = \sum_i a_i \sum_j B_{ij} .,$$

and, from the definition of A_i, $A_i - \sum_j B_{ij} = \sum_j Q_{ij}$. It follows that

$$\text{Adj. S.S. } (A) = \textstyle\sum_i a_i \sum_j Q_{ij} \tag{5.20}$$

and

$$\textstyle\sum_j Q_{ij} = n\lambda_2 v a_i / k, \tag{5.21}$$

the latter result obtained from (4.4), (5.2) to (5.6), and algebraic reduction based on (vii) of Section 2. The final form for Adj. S.S. (A) given in (5.8) is now evident and the degrees of freedom are $(m - 1)$, since $(m - 1)$ additional restrictions were required to reduce Ω to ω_A. Adj. S.S. (C) and Adj. S.S. (AC) are obtained in much the same way.

It is of interest in some applications to have the variances of $(a_i - a_{i'})$, $i \neq i'$, and of $(c_j - c_{j'}), j \neq j'$. These variances are most easily obtained from the forms of the multipliers of $\sum_i \bar{t}_{i\cdot}^2 = \sum_i a_i^2$ and of $\sum_j \bar{t}_{\cdot j}^2 = \sum_j c_j^2$ in the analysis of variance of Table 1. It follows that

$$V(a_i - a_{i'}) = 2k\sigma^2 / n\lambda_2 v, \qquad i \neq i', \tag{5.22}$$

and

$$V(c_j - c_{j'}) = 2k\sigma^2 / m(\lambda_1 + rk - r), \qquad j \neq j'. \tag{5.23}$$

The error mean square of the analysis of variance is used to estimate σ^2 and consequently the variances of (5.22) and (5.23). Alternate derivations of (5.22) and (5.23) may be obtained through the forms (5.11) and (5.12) given the variances and covariances of the t_{ij}. Considerable algebra is involved in the derivation of these variances and covariances, and we do not include it here. It may, however, be useful to have these results and we now state without proof that

$$V(t_{ij}) = k\sigma^2 \left[\frac{(n-1)}{n(\lambda_1 + rk - r)} + \frac{(m-1)}{mn\lambda_2 v} \right], \tag{5.24}$$

$$\text{Cov } (t_{ij} t_{ij'}) = k\sigma^2 \left[\frac{m-1}{mn\lambda_2 v} - \frac{1}{n(\lambda_1 + rk - r)} \right], \qquad j \neq j', \tag{5.25}$$

and

$$\text{Cov } (t_{ij} t_{i'j'}) = -k\sigma^2 / mn\lambda_2 v, \qquad i \neq i'. \tag{5.26}$$

Efficiencies of factorial contrasts may be obtained in the same way as E_1 and E_2 in (4.11) and (4.12). The variances corresponding to (5.22) and (5.23) respectively for a randomized block design are $2\sigma^2/rn$ and $2\sigma^2/rm$, on the assumption again of equal experimental errors for the complete and incomplete block designs. The efficiency for contrasts among A-factor effects is

$$E_A = \lambda_2 v/rk \tag{5.27}$$

and the efficiency for contrasts among C-factor effects is

$$E_C = (\lambda_1 + rk - r)/rk. \tag{5.28}$$

The variance for an interaction contrast in the group divisible design is

$$(\lambda_1 + rk - r)\sigma^2/k$$

from Table 1 and is σ^2/r for the randomized block design. Consequently, the efficiency for an AC-interaction contrast is also

$$E_{AC} = (\lambda_1 + rk - r)/rk. \tag{5.29}$$

Note that $E_C = E_{AC} = E_1$. The two-associate class group divisible designs have three subclasses as noted earlier. For the singular subclass, $\lambda_1 = r$ and $E_C = E_{AC} = 1$; for the semi-regular subclass, $\lambda_2 v = rk$ and $E_A = 1$. In the next section we discuss individual comparisons and multifactor factorials. We now note, somewhat in advance, that all individual comparisons and subfactor effects of the A-factor have the efficiency E_A, those of the C-factor have efficiency E_C, and those of the AC-interaction have efficiency E_{AC}.

6. Individual comparisons and multi-factor factorials. Individual or single-degree-of-freedom comparisons are possible in much the usual way.

Let ξ be an $(m-1)$ by m orthogonal matrix and η, an $(n-1)$ by n orthogonal matrix used to transform the α's and γ's respectively. Contrasts on A-factor effects would be

$$\xi_u = \textstyle\sum_i \xi_{iu}\alpha_i\,, \qquad u = 1, \cdots, (m-1), \tag{6.1}$$

and on C-factor effects,

$$\eta_v = \textstyle\sum_j \eta_{vj}\gamma_j\,, \qquad v = 1, \cdots, (n-1). \tag{6.2}$$

To test the hypothesis that $\xi_u = 0$, we form the contrast

$$I_u = \textstyle\sum_i \xi_{iu}\bar{l}_{i\cdot} = \sum_i\sum_j \xi_{iu}l_{ij}/n \tag{6.3}$$

and

$$\begin{aligned}\text{Adj. S.S. } (I_u) &= n\lambda_2 v(\textstyle\sum_i \xi_{iu}\bar{l}_{i\cdot})^2/k\sum_i \xi_{iu}^2 \\ &= \lambda_2 v(\textstyle\sum_i\sum_j \xi_{iu}l_{ij})^2/k\sum_i\sum_j \xi_{iu}^2\,.\end{aligned} \tag{6.4}$$

Similarly, to test the hypothesis that $\eta_v = 0$, we form the contrast

$$J_v = \textstyle\sum_j \eta_{vj}\bar{l}_{\cdot j} = \sum_i\sum_j \eta_{vj}l_{ij}/m \tag{6.5}$$

and

$$\begin{aligned}\text{Adj. S.S. } (J_v) &= m(\lambda_1 + rk - r)(\textstyle\sum_j \eta_{vj}\bar{t}_{\cdot j})^2/k\textstyle\sum_j \eta_{vj}^2 \\ &= (\lambda_1 + rk - r)(\textstyle\sum_i\sum_j \eta_{vj}t_{ij})^2/k\textstyle\sum_i\sum_j \eta_{vj}^2 .\end{aligned} \tag{6.6}$$

The contrast for interaction of ξ_u and η_v is

$$(\xi\eta)_{uv} = \textstyle\sum_i\sum_j \xi_{iu}\eta_{vj}\delta_{ij} . \tag{6.7}$$

The hypothesis, $(\xi\eta)_{uv} = 0$, is tested through use of the contrast

$$(IJ)_{uv} = \textstyle\sum_i\sum_j \xi_{iu}\eta_{vj}t_{ij} \tag{6.8}$$

and

$$\begin{aligned}&\text{Adj. S.S. } (IJ)_{uv} = (\lambda_1 + rk - r) \\ &\cdot(\textstyle\sum_i\sum_j \xi_{iu}\eta_{vj}t_{ij})^2/k\textstyle\sum_i\sum_j (\xi_{iu}\eta_{vj})^2 .\end{aligned} \tag{6.9}$$

Cochran's theorem [11] is sufficient to demonstrate the independence of all adjusted sums of squares, Adj. S.S. (I_u), $u = 1, \cdots, (m-1)$, Adj. S.S. (J_v), $v = 1, \cdots, (n-1)$, and Adj. S.S. $(IJ)_{uv}$, $u = 1, \cdots, (m-1)$, $v = 1, \cdots, (n-1)$, and that they are appropriate for use in analysis of variance. Each has one degree of freedom and F-tests are effected using the error mean square of Table 1.

Special definition of the matrices ξ and η permits the use of special contrasts. For example, rows of ξ and η may be defined such that contrasts on A-factor and C-factor effects measure trends (linear, quadratic, cubic, $\cdots$) over the factor levels.

Suppose the A-factor has levels which themselves are factorial combinations of other factors. Let there be p such factors, $A_1, \cdots, A_p$, with levels $m_1, \cdots, m_p$. It is only required that $m = \prod_{i=1}^{p} m_i$. Then ξ may be chosen in the obvious way so that the contrasts defined may be grouped to obtain main-effect and interaction comparisons for the subfactors of A. The corresponding adjusted sums of squares, each with one degree of freedom, may be grouped if desired to obtain Adj. S.S. (A_1) with $(m_1 - 1)$ degrees of freedom, Adj. S.S. (A_2) with $(m_2 - 1)$ degrees of freedom, Adj. S.S. for interaction of A_1 and A_2 with $(m_1 - 1)(m_2 - 1)$ degrees of freedom, etc. Alternately, these sums of squares may be computed by forming the usual two-way, three-way, etc., tables of values of $\bar{t}_{i\cdot}$ and effecting the computation as though they were observations in a single replication on factorial treatment combinations only finally multiplying the resulting sums of squares by the coefficient $n\lambda_2 v/k$ of (6.4). Similarly the C-factor may consist of factorial combinations of q factors, $C_1, \cdots, C_q$, with levels $n_1, \cdots, n_q$ such that $\prod_{j=1}^{q} n_j = n$ and appropriate contrasts and adjusted sums of squares may be obtained with proper selection of the rows of η. When ξ and η have been defined, the corresponding contrasts for interaction of A-factor and C-factor contrasts follow immediately. These in turn yield adjusted sums of squares that may be grouped to yield sums of squares for interaction of A_1 and C_1, A_1, A_2, and C_1, etc.

Now we have shown how multi-factor factorials may be used in two-associate class group divisible designs. It is also evident that fractional factorials may be used. The levels of the A-factor may be designated to be m treatment combinations of a fractional factorial which is a fraction of a full factorial with, say, hm treatment combinations, h an integer. The levels of the A-factor would then form a $(1/h)$-th fraction of the full factorial. Similarly, the n levels of the C-factor might be a fraction of a second set of factorial treatment combinations. Analysis of the resulting fractional factorial experiment would again depend only on proper specification of ξ and η. We would have r replications of a fractional factorial in the experiment. This may be a very useful system when it is necessary to use small incomplete blocks in a study.

7. Remarks. We have shown how factorials may be incorporated in group divisible partially balanced incomplete block designs with two associate classes. The factorial treatment combinations were so matched with treatments in the rectangular association schemes for these designs as to yield quite simple analyses. Other correspondences between factorial treatments and the treatments of the basic designs may be possible, but we would expect that they would result in considerably more complex analyses and in lack of orthogonality among the factorial comparisons. The problem of the recovery of inter-block information is being considered.

The group divisible subclass of the two-associate class of partially balanced incomplete block designs is only one of five subclasses given in [7]. The others listed are Simple, Triangular, Latin Square Type, and Cyclic, and comprise only a minor percentage of the designs listed in the reference. In particular, many designs of the Simple and Cyclic subclasses have values of v which are prime numbers and are not therefore suitable for factorials. Factorials have been developed in Simple and Triangular designs for special cases, but a general development has not been found.

In the view of the authors, important applications of these factorial incomplete block designs should be forthcoming. They should be useful in large animal experimentation where litter sizes sharply limit the amounts of homogeneous experimental material available. In taste testing, fatigue and other factors limit the number of samples that can be considered at a session, and these designs have applications there. In industrial experimentation, it may not be possible to make many observations while normal production is interrupted, and again use of incomplete blocks may be desirable. Some numerical examples on the uses of these designs are being prepared for an applied paper [12].

Marvin Zelen [13] did some preliminary work on the use of factorials in incomplete block designs (and subsequently has obtained additional results independent of us.) While the formulation and presentation given here are our own, we wish to acknowledge his cooperation through helpful discussions when this research was initiated.

REFERENCES

[1] R. C. Bose, and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[2] R. C. Bose, S. S. Shrikhande, and K. N. Bhattacharya, "On the construction of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 167–195.

[3] Bose, R. C. and Shimamoto, T., "Classification and analysis of partially balanced designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–190.

[4] Oscar Kempthorne, *The Design and Analysis of Experiments*, John Wiley and Sons, Inc., 1952.

[5] C. Y. Kramer, and R. A. Bradley, "$k(k-q)$ factorial treatments in near balance incomplete block designs," Progress Report No. 3 on Statistical Techniques for Sensory Testing of Processed Foods, Contract Report with U. S. D. A.

[6] Boyd Harshbarger, "The 2^3 Factorial in a Latinized Rectangular Lattice Design," *Technical Report No. 1 (1954) to Office of Ordnance Research, OOR Project No. 1166.*

[7] R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, "Tables of Partially Balanced Designs with Two Associate Classes," *Tech. Bul. No. 107 (1954), North Carolina Agricultural Experiment Station.*

[8] Boyd Harshbarger, "Near Balance Rectangular Lattices," *Virginia Jour. of Sci.*, Vol. II (new series) (1951), pp. 13–27.

[9] Boyd Harshbarger, "Latinized Rectangular Lattices," *Biometrics*, Vol. 8 (1952), pp. 73–84.

[10] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, 1950.

[11] W. G. Cochran, "The distribution of quadratic forms in a normal system with applications to the analysis of covariance," *Proc. Cambridge Philos. Soc.*, Vol. 30 (1934), pp. 178–191.

[12] C. Y. Kramer, and R. A. Bradley, "Examples of intra-block analysis for factorials in two-associate class group divisible designs" (in preparation).

[13] Marvin Zelen, "The use of incomplete block designs for factorial experiments" (unpublished manuscript presented to Amer. Stat. Assn. at Montreal, Canada, Sept. 9–10, 1954).

STATISTICAL PROPERTIES OF INVERSE GAUSSIAN DISTRIBUTIONS. I.

By M. C. K. Tweedie[1]

Virginia Agricultural Experiment Station, Virginia Polytechnic Institute

0. Summary. A report is presented on some statistical properties of the family of probability density functions

$$\exp\left[-\lambda(x-\mu)^2/2\mu^2 x\right][\lambda/2\pi x^3]^{1/2}$$

for a variate x and parameters μ and λ, with x, μ, λ each confined to $(0, \infty)$. The expectation of x is μ, while λ is a measure of relative precision. The chief result is that the ml estimators of μ and λ have stochastically independent distributions, and are of a nature which permits of the construction of an analogue of the analysis of variance for nested classifications. The ml estimator of μ is the sample mean, and for a fixed sample size n its distribution is of the same family as x, with the same μ but with λ replaced by λn. The distribution of the ml estimator of the reciprocal of λ is of the chi-square type. The probability distribution of $1/x$, and the estimation of certain functions of the parameters in heterogeneous data, are also considered.

1. Introduction. The name "Inverse Gaussian" has been suggested [1] for the members of a certain family of continuous probability density functions in which the variate takes positive values only. The family is generated by varying the values of two real positive parameters, which may be any independent pair from α, λ, μ, ϕ, where $\frac{1}{2}\alpha^2 = \mu = \lambda/\phi$. The density function for the positive values of the variate may accordingly be written in the forms

$$\text{(1a)}\quad f_1(x;\alpha,\lambda) = \exp\left\{-\alpha\lambda x + \lambda(2\alpha)^{1/2} - \lambda/2x\right\}[\lambda/2\pi x^3]^{1/2}.$$

$$\text{(1b)}\quad f_2(x;\mu,\lambda) = \exp\left\{-\lambda(x-\mu)^2/2\mu^2 x\right\}[\lambda/2\pi x^3]^{1/2},$$

$$\text{(1c)}\quad f_3(x;\mu,\phi) = \exp\left\{-\frac{\phi x}{2\mu} + \phi - \frac{\mu\phi}{2x}\right\}[\mu\phi/2\pi x^3]^{1/2},$$

$$\text{(1d)}\quad f_4(x;\phi,\lambda) = \exp\left\{-\frac{\phi^2 x}{2\lambda} + \phi - \frac{\lambda}{2x}\right\}[\lambda/2\pi x^3]^{1/2}.$$

Each of these forms is convenient or suggestive for some purpose.

The relationships

$$\text{(2)}\qquad f_2(x;\mu,\lambda) = \mu^{-1}f_3(x/\mu;1,\phi) = \lambda^{-1}f_4(x/\lambda;\phi,1)$$

Received April 9, 1956; revised July 3, 1956.

[1] Present address: Statistical Laboratory, Department of Mathematics, The University, Manchester, England.

are useful in computing numerical values of the probability density. The cumulative probability function depends essentially on only two variables, which might be chosen to be x/μ and ϕ. The case $\mu = 1$ could therefore be adopted as a standard form. Curves of the density functions for $\lambda = \phi = \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16, 32$, with $\mu = 1$, are shown in Fig. 1. In some physical applications it is more natural to hold λ constant, and Fig. 2 shows the density curves for $\lambda = 1$ with $\mu = 4, 1$, and $\frac{1}{4}$, i.e., for $\phi = \frac{1}{4}, 1$, and 4 respectively.

Since it will be found useful to consider also some functions of the same algebraic form but with complex values for some of the parameters, it may be noted that the integrals of functions such as (1) over the interval $(0, \infty)$ can be shown to be unity, provided that the real parts of λ and of the mutually equal quantities $\alpha\lambda$ and $\frac{1}{2}\lambda\mu^{-2}$ are positive. For reference we reproduce an equation for a modified Bessel function of the second kind,

$$K_{\pm\nu}(z) = \tfrac{1}{2}(\tfrac{1}{2}z)^{\nu} \int_0^{\infty} \exp\left\{-t - \frac{z^2}{4t}\right\} \frac{dt}{t^{\nu+1}}, \tag{3}$$

given by Watson ([2], p. 183), under the condition that the real part of z^2 is positive, together with the result

$$K_{\pm 1/2}(z) = e^{-z}(\pi/2z)^{1/2}, \tag{4}$$

also given by Watson ([2], p. 80).

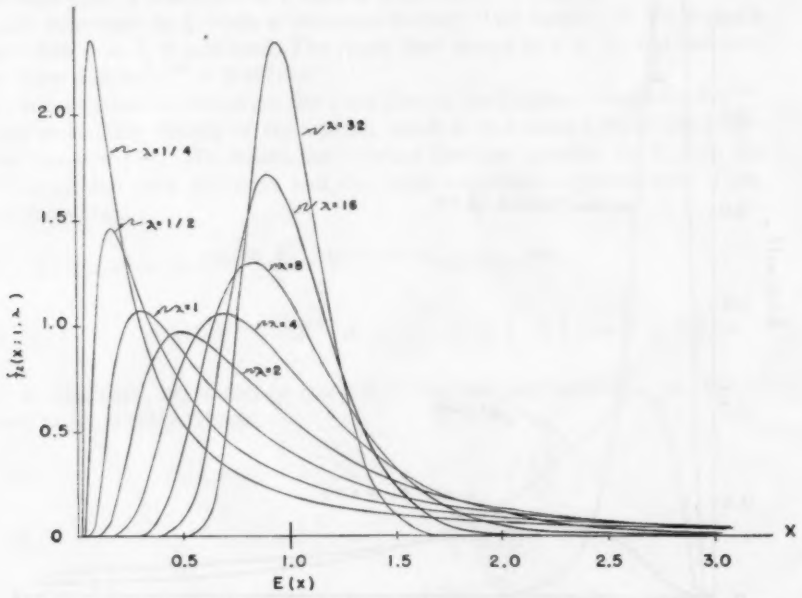


FIG. 1. Probability density curves for an Inverse Gaussian variate with $\mu = 1$ for 8 values of λ or ϕ.

The Inverse Gaussian family of distributions arises in a problem of Brownian motion (cf. [1], [3]), though then a further parameter appears in the physical formulation. The numerical value of this parameter can however normally be regarded as known, and it merely modifies the values of the parameters given in the expressions (1) above. Both λ and μ are of the same physical dimensions as the random variable x itself. A change of scale of x, such as may be due to a change in measuring unit or, approximately, to changes in temperature or some other factor, produces another member of the family, in which λ and μ have been multiplied by the same factor as x. The ratio ϕ is invariant under such a change.

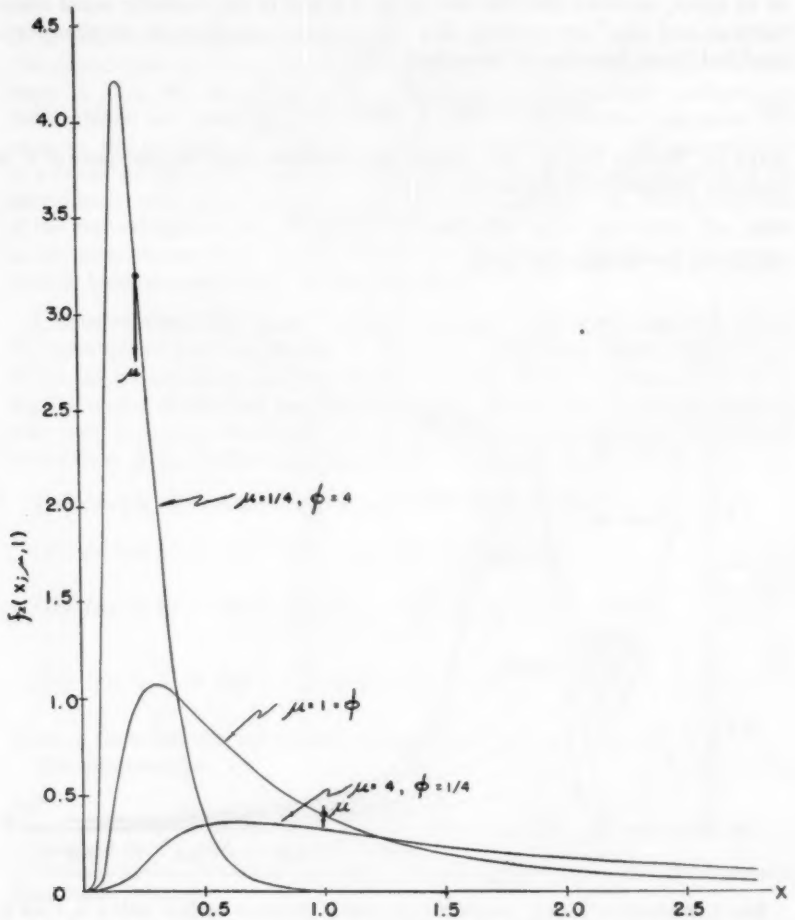


FIG. 2. Probability density curves for an Inverse Gaussian variate with $\lambda = 1$ for 3 values of μ or ϕ.

The same family appears also as a limiting form for the distribution of the final sample size in a special case of Wald's sequential likelihood ratio test [4]. Some properties of this family were studied in a degree thesis [5], where the Brownian motion problem was found to have an important part in the interpretation of some experimental work. The present paper establishes some of those exact properties in a more formal way, though using essentially the same methods as in the thesis. Some new results are included, and some further ones will be given in another paper. Not all these results are of quantitative importance in the original physical problem, and those which are not are presented here for their theoretical interest. The formulae (1) will be regarded as given, in that no derivations will be offered here. The uniqueness of certain Laplace transforms will be an important factor in some of the proofs. The form (1a) is of the kind adopted in an earlier published paper [6], in which similar methods were used.

2. Basic characteristics.

The shape of the distribution depends on ϕ only. The distribution is unimodal, with its mode at

$$x_{\text{mode}} = \mu\left\{\left(1 + \frac{9}{4\phi^2}\right)^{1/2} - \frac{3}{2\phi}\right\}. \tag{5}$$

The ratio x_{mode}/μ converges to 1 when ϕ is increased to infinity; while the ratio x_{mode}/λ converges to $\frac{1}{3}$ when ϕ decreases to zero. The density at the mode is least when $\phi = 2$, if μ is fixed. The mode then occurs at $x = \frac{1}{2}\mu$ and the density there is $[8/\pi\mu^2 e]^{1/2} = 0.96788\mu^{-1}$.

It is convenient to introduce the logarithm of the Laplace transform $E(e^{-tx})$ of the probability density of the variate, which is in a sense a cumulant-generating function (cgf). We denote the relevant function operator by L, with the variate symbol as a subscript and the other variables in parentheses. Thus, from form (1a),

$$L_x(t; \mu, \lambda) = \ln e^{\lambda(2\alpha)^{1/2}} \int_0^\infty e^{-(\alpha+t/\lambda)\lambda x - \lambda/2x}\, dx[\lambda/2\pi x^3]^{1/2} \tag{6}$$

$$= \lambda(2\alpha)^{1/2} - \lambda 2^{1/2}\left(\alpha + \frac{t}{\lambda}\right)^{1/2} + \ln \int_0^\infty f_1\left(x; \alpha + \frac{t}{\lambda}, \lambda\right) dx. \tag{7}$$

If t is imaginary, or, if real or complex, if its real part exceeds $-\alpha\lambda$, the integral in (7) is unity. Hence

$$L_x(t; \mu, \lambda) = \lambda\left\{(2\alpha)^{1/2} - 2^{1/2}\left(\alpha + \frac{t}{\lambda}\right)^{1/2}\right\} \tag{8a}$$

$$= \frac{\lambda}{\mu}\left\{1 - \left(1 + \frac{2\mu^2 t}{\lambda}\right)^{1/2}\right\} \tag{8b}$$

$$= \phi\left\{1 - \left(1 + \frac{2\mu t}{\phi}\right)^{1/2}\right\} \tag{8c}$$

$$= \phi\left\{1 - \left(1 + \frac{2\lambda t}{\phi^2}\right)^{1/2}\right\}. \tag{8d}$$

This cgf. is unique to the density function (1).

The cumulants can be obtained from the power series expansion of $L_s(t; \mu, \lambda)$. They are:

$$\kappa_1 = \mu = \lambda\phi^{-1}, \qquad \kappa_2 = \mu^3\lambda^{-1} = \lambda^2\phi^{-3},$$

$$\kappa_3 = 3\mu^5\lambda^{-2} = 3\lambda^3\phi^{-5}, \qquad \kappa_4 = 15\mu^7\lambda^{-3} = 15\lambda^4\phi^{-7},$$

and, in general, when $r \geqq 2$,

$$\begin{aligned}\kappa_r &= 1\cdot 3\cdot 5 \cdots (2r-3)\mu^{2r-1}\lambda^{1-r} \\ &= \lambda^r(2r-3)!/\phi^{2r-1}2^{r-2}(r-2)!.\end{aligned} \tag{9}$$

Thus μ is the population mean and is primarily a measure of location, while λ is an inverse measure of relative dispersion, being the ratio of κ_1^3 to κ_2, or

$$\frac{1}{\lambda} = \frac{\kappa_2}{\kappa_1^3}. \tag{10}$$

Also, $\phi = \kappa_1^2/\kappa_2$. The Fisherian shape coefficients, or standardized cumulants, are

$$\begin{aligned}\gamma_1 &= \kappa_3\kappa_2^{-3/2} = 3\phi^{-1/2}, \qquad \gamma_2 = \kappa_4\kappa_2^{-2} = 15\phi^{-1}, \\ \gamma_r &= \kappa_{r+2}\kappa_2^{-r/2-1} = 1\cdot 3\cdot 5 \cdots (2r+1)\phi^{-r/2}.\end{aligned} \tag{11}$$

The fractional coefficient of variation is $\gamma_{-1}^{-1} = \kappa_1^{-1}\kappa_2^{1/2} = \phi^{-1/2}$, so that $\gamma_1 = 3\gamma_{-1}^{-1}$, $\gamma_2 = 15\gamma_{-1}^{-2}$, and so on. Evidently the distribution becomes more and more nearly normal when ϕ is increased. This parameter ϕ might be called the normality parameter or the shape parameter.

In the probability density curves shown in Fig. 1, γ_1 ranges from 6 down to 0.53, and γ_2 ranges from 60 down to 0.47. The approach to normality in the neighbourhood of $x = \mu$ is evident from these curves. However, some important aspects of the distributions, such as the standardized cumulants, depend primarily on the behaviour of the functions at very large values of the variate, whereas the diagrams are necessarily bounded.

The positive integral moments about zero are obtainable either from (9) or by direct integration, using (3) and a further result given by Watson ([2], Eq. (12), p. 80). They are:

$$\begin{aligned}\mu_1' &= \mu, \qquad \mu_2' = \mu^2 + \mu^3\lambda^{-1}, \\ \mu_2' &= \mu^3 + 3\mu^4\lambda^{-1} + 3\mu^5\lambda^{-2}, \\ \mu_4' &= \mu^4 + 6\mu^5\lambda^{-1} + 15\mu^6\lambda^{-2} + 15\mu^7\lambda^{-3},\end{aligned} \tag{12}$$

$$\mu_r' = \mu^r K_{r-1/2}(\phi)K_{1/2}(\phi) = \mu^r \sum_{s=0}^{r-1} \frac{(r-1+s)!}{s!(r-1-s)!(2\phi)^s}.$$

The negative integral moments are given in (33).

It follows from the form of (8) that the distribution of the arithmetic mean of a fixed number n of independent values from (1) is a member of the same family, with the same α and μ, but with λ replaced by λn and ϕ replaced by ϕn. More generally, suppose that we have a set of populations in which μ_i and λ_i are the values of the parameters in the i-th population, and that, although the values of these parameters are unknown, the values of $a_i = C\mu_i^{-2}\lambda_i$ are known, C being a constant whose value is not necessarily known. The distribution of the linear function $\sum_{i=1}^{n}(a_i x_i)$ is then of the same form as (1), with $\mu = C\sum_{i=1}^{n}\phi_i$, $\lambda = C(\sum_{i=1}^{n}\phi_i)^2$, $\phi = \sum_{i=1}^{n}\phi_i$. Because of this additive property of the normality parameter, the linear function will have a more nearly normal distribution than any of its components.

3. Estimation of parameters. Suppose that x_i is an observation on a distribution of the form (1b), with parameter values μ and λ_i, where $\lambda_i = \lambda_0 w_i$ for $i = 1$ to N, w_i being positive and known, but neither of the common values of μ and λ_0 being known. For example, x_i might be the arithmetic mean of w_i values from a distribution with parameter values μ and λ_0. With these N pairs of values of x_i and w_i as data, the estimates of μ and λ_0 which jointly maximize the likelihood function are given by

$$\hat{\mu} = x_{\cdot} = \sum_{i=1}^{N}(w_i x_i) \Big/ \sum_{i=1}^{N}(w_i), \tag{13}$$

$$\frac{1}{\hat{\lambda}_0} = \frac{1}{N}\sum_{i=1}^{N} w_i\left(\frac{1}{x_i} - \frac{1}{x_{\cdot}}\right). \tag{14}$$

These estimates can never be negative so long as the observations are necessarily non-negative. For (14) this follows from the concavity of the function x^{-1}. With every w_i equal to unity, these estimates were given by Schrödinger [3], who called them "wahrscheinlichste."

The Inverse Gaussian family is one for which the weighted sample mean $x_{\cdot} = \hat{\mu}$ (13) is a sufficient statistic (in Fisher's sense) for estimating the common population mean μ. Further, the cumulant-generating function of $\hat{\mu}$, with fixed values of μ, λ_0, N and the weights $w_1, \cdots w_N$, differs from (8) only in that λ becomes $\lambda_0\sum_{i=1}^{N} w_i$. (To see this, take $C = \mu^2/\lambda_0\sum_{i=1}^{N} w_i$ in the result at the end of Section 2.) The probability density function of $\hat{\mu}$ therefore is

$$f_2(\hat{\mu};\ \mu,\ \lambda_0\textstyle\sum_{i=1}^{N} w_i), \qquad 0 < \hat{\mu} < \infty. \tag{15}$$

In the terminology of a previous paper [6], the family (1) is a Laplacian one with α as primary parameter and λ as secondary parameter. Hence $\hat{\mu}$ has a Laplacian form of probability density function. This enables the conditional moments and cumulants and other properties of other statistics, with a fixed value of $\hat{\mu}$, to be found by using the uniqueness of the Laplace transforms which appear in their mathematical formulations. A number of exact results have been found for the Inverse Gaussian distributions in this way, and we shall now proceed to develop one of the more surprising of them.

4. Distribution of the ml estimator of the secondary parameter. With the same data and the same fixed quantities as were introduced in Section 3 in discussing the distribution of the maximum likelihood (ml) estimator of μ, the Laplace transform of the probability density function of $1/\hat{\lambda}_0$ is

$$E(e^{-t/\hat{\lambda}_0}) = \int_{\text{all } x_i>0} \cdots \int e^{-t/\hat{\lambda}_0} \prod_{i=1}^{N} f_2(X_i; \mu, \lambda_0 w_i)\, dX_i. \tag{16}$$

This certainly exists when the real part of t is not negative. On substituting from (1) and (15) and writing $\sum_{i=1}^{N} w_i = W$ for brevity, we get

$$E(e^{-t/\hat{\lambda}_0} \mid \mu, \lambda_0, w_1, \cdots w_N, N) = \int_{\hat{\mu}=0}^{\hat{\mu}=\infty} f_2(\hat{\mu}; \mu, \lambda_0 W) \int_{\hat{\mu} \text{ constant}} \cdots \int e^{-(t+\lambda_0 N/2)/\hat{\lambda}_0} \cdot \frac{\hat{\mu}^{3/2}}{W^{1/2}} \left(\frac{\lambda_0}{2\pi}\right)^{(N-1)/2} \prod_{i=1}^{N} \frac{w_i^{1/2}\, dX_i}{X_i^{3/2}}. \tag{17}$$

The multiple integral in the final integrand on the right of (17) does not contain μ or α. From the Laplacian form of $f_2(\hat{\mu}; \mu, \lambda_0 W)$ and the uniqueness of the Laplace transforms to which it gives rise, it follows (cf. [6]) that the partial derivative, with respect to $\hat{\mu}$, of this multiple integral is equal to the Laplace transform of the conditional density of $1/\hat{\lambda}_0$. This statement may be justified by reference either to Lerch's theorem ([7], p. 52; [8], p. 61) or to an equally applicable set of theorems (cf. [9], p. 38). The proof can legitimately involve α taking complex values with positive real parts. Therefore

$$\frac{\partial}{\partial \hat{\mu}} \int_{\hat{\mu} \text{ constant}} \cdots \int e^{-(t+\lambda_0 N/2)/\hat{\lambda}_0} \hat{\mu}^{3/2} W^{-1/2} (\lambda_0/2\pi)^{(N-1)/2} \prod_{i=1}^{N} (X_i^{-3/2} w_i^{1/2}\, dX_i) = E(e^{-t/\lambda_0} \mid \hat{\mu}; \lambda_0, w_1, \cdots w_N, N), \tag{18}$$

which is the conditional moment-generating function of $1/\hat{\lambda}_0$ with $\hat{\mu}$ fixed. To evaluate this integral, first take $t = 0$, which gives

$$\frac{\partial}{\partial \hat{\mu}} \int_{\hat{\mu} \text{ constant}} \cdots \int e^{-\lambda_0 N/2\hat{\lambda}_0} \hat{\mu}^{3/2} \prod_{i=1}^{N} (X_i^{-3/2}\, dX_i) = \frac{W^{1/2}}{\prod_{i=1}^{N} (w_i^{1/2})} \left(\frac{2\pi}{\lambda_0}\right)^{(N-1)/2}. \tag{19}$$

By substituting $\lambda_0 + 2t/N$ for λ_0 on both sides of (19), the left-hand side of (18) is found almost immediately to be

$$E(e^{-t/\hat{\lambda}_0} \mid \hat{\mu}; \lambda_0, w_1, \cdots w_N, N) = \left(1 + \frac{2t}{\lambda_0 N}\right)^{(N-1)/2}. \tag{20}$$

This is a Laplace transform of a density function of the chi-square type, with N-1 degrees of freedom. In fact, it shows that

$$\frac{1}{\hat{\lambda}_0} = \frac{\chi^2_{(N-1 \text{ d.f.})}}{\lambda_0 N} \tag{21}$$

and thus that

$$\sum_{i=1}^{N} w_i \left(\frac{1}{x_i} - \frac{1}{x_.}\right) = \frac{\chi^2_{(N-1 \text{ d.f.})}}{\lambda_0}. \tag{22}$$

This result can be used to obtain confidence intervals for λ_0. By substituting for the probability elements of $\hat{\mu}$ and $\hat{\lambda}_0$ in the joint probability element of the N observations, it can be shown also that $\hat{\mu}$ and $\hat{\lambda}_0$ are jointly sufficient estimators of μ and λ_0. Further,

$$\frac{1}{N-1} \sum_{i=1}^{N} w_i \left(\frac{1}{x_i} - \frac{1}{x_.}\right) \tag{23}$$

is an unbiased estimator of $1/\lambda_0$. Its distribution is of exactly the same type as that of the usual unbiased quadratic estimator of the variance of a Gaussian distribution, although it cannot be expressed precisely as a sum of squares of Gaussian variates with zero means. The conditional distribution of (23) is necessarily independent of μ, because of the sufficiency of $\hat{\mu}$, but it is also independent of $\hat{\mu}$, thus affording the possibility of an analogue of the analysis of variance, using the existing tables of χ^2 and F for significance tests and so forth. In the Brownian motion problem μ and λ are concerned with rather different physical properties of the experimental system, which do however occur together in some physical formulae. It is therefore convenient that estimators have been found which are both independently distributed and jointly sufficient.

The statistic (14) appears also in the likelihood ratio test of the hypothesis that the population means are equal against the alternative hypothesis that they may have any values independently of one another, the values for the means not being specified, while the value of the secondary parameter is supposed known. The logarithm of the ratio of the maximum likelihoods under these two hypotheses is $-\lambda_0 N/2\hat{\lambda}_0$, so that the result (21) is a case where the well-known approximate general result ([9], p. 151) holds exactly. Moreover, the statistic $\hat{\lambda}_0$ depends essentially on the difference between the ml estimator of the reciprocal of the hypothetically common value for μ and the ml estimator of the weighted mean of the reciprocals of the means under the assumption of their complete independence. Both these considerations indicate that $\hat{\lambda}_0$ will tend to be increased by real differences between the population means, and that it therefore measures the combined effect of the dispersion in homogeneous samples and the heterogeneity of the means.

5. An analogue of the analysis of variance for nested classifications. The algebraic aspect of the analysis of variance for nested classifications may be generalized, for two classifications (which is a sufficiently general case), to

$$\begin{aligned} &\textstyle\sum_{i=1}^{N} \sum_{j=1}^{n_i} \{\psi(x_{ij}) - \psi(x_{..})\} \\ &= \textstyle\sum_{i=1}^{N} \sum_{j=1}^{n_i} \{\psi(x_{ij}) - \psi(x_{i.})\} + \sum_{i=1}^{N} n_i \{\psi(x_{i.}) - \psi(x_{..})\}. \end{aligned} \tag{24}$$

Here x_{ij} is the observed value of the variate in the j-th subclass of the i-th major class, and n_i is the number of subclasses in this major class, while there are N major classes. The values of $x_{i.}$ and $x_{..}$ are means of some kind of the values of x_{ij}, and $\psi(x)$ is some suitable function of x. The sums might be regarded as depending on the differences between different kinds of means, in that they could be rewritten as

$$\left\{\frac{1}{n_.}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\psi(x_{ij}) - \psi(x_{..})\right\}$$

$$= \frac{1}{n_.}\sum_{i=1}^{N} n_i \left\{\frac{\sum_{j=1}^{n_i}\psi(x_{ij})}{n_i} - \psi(x_{i.})\right\} + \left\{\frac{1}{n_.}\sum_{i=1}^{N}[n_i\psi(x_{i.})] - \psi(x_{..})\right\}, \tag{25}$$

where $n_. = \sum_{i=1}^{N} n_i$. That is to say, if we temporarily use M to stand for the operation of taking the relevant mean involved in $x_{i.}$ and $x_{..}$, and A to stand for the operation of taking the weighted arithmetic mean, the identity (25) can be written

$$A_i A_j \psi - \psi M_i M_j = A_i(A_j\psi - \psi M_j) + (A_i\psi - \psi M_i)M_j, \tag{26}$$

operating on x_{ij}. Certain restrictions on ψ and M are necessary to ensure that these differences, which are essentially measures of dispersion, shall never change sign. In the analysis of variance, the means entailed by M are arithmetical, while $\psi(x) = x^2$. If the variates have independent Gaussian distributions with both means and variances equal within any major group, the two major sums on the right side of (24) have independent distributions. This independence does not generally occur in other circumstances, but it is available to some extent with the Inverse Gaussian distribution. For this, according to Section 4, we again take the means M to be arithmetical, but take $\psi(x) = x^{-1}$.

From the results obtained in Section 4, we see that the statistic

$$\sum_{j=1}^{n_i}\left(\frac{1}{x_{ij}} - \frac{1}{x_{i.}}\right) = \frac{n_i}{\lambda_i} \tag{27}$$

is distributed as χ^2/λ_i with $n_i - 1$ degrees of freedom and independently of $x_{i.}$, $= \sum_{j=1}^{n_i}(x_{ij})/n_i$. Hence

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{1}{x_{ij}} - \frac{1}{x_{i.}}\right)$$

is distributed as χ^2/λ with $n_. - N$ degrees of freedom and independently of the values of $x_{i.}$. (This distribution would remain true even if the expectations $E(x_{i.})$ varied with i.) In particular, if $x_{..} = \sum_{i=1}^{N}(n_i x_{i.})/n_i$, that double sum is distributed independently of

$$\sum_{i=1}^{N} n_i\left(\frac{1}{x_{i.}} - \frac{1}{x_{..}}\right),$$

which is itself distributed as χ^2/λ with $(N - 1)$ degrees of freedom.

The algebraic identity (24) thus becomes

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{1}{x_{ij}}-\frac{1}{x_{..}}\right)=\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{1}{x_{ij}}-\frac{1}{x_{i.}}\right)+\sum_{i=1}^{N}n_i\left(\frac{1}{x_{i.}}-\frac{1}{x_{..}}\right). \tag{28}$$

If all the observations come independently from the same Inverse Gaussian distribution, the three major sums in (28) are each distributed as χ^2/λ, the chi-squares having respectively $n_. - 1$, $n_. - N$, and $N - 1$ degrees of freedom. Thus $1/\lambda$ can be estimated by dividing any of the three sums in (28) by the corresponding number of degrees of freedom. The two sums on the right of (28) have independent distributions and therefore their ratio is distributed as

$$(N-1)F/(n_.-N),$$

where F has $N - 1$ and $n_. - N$ degrees of freedom. Hence the analogy with the analysis of variance is very close. For example, a significance test of the differences between the N values of $x_{i.}$ may be made by using the first major sum on the right of (28) as the analogue of the sum of squares for error. Some illustrations of the use of these formulae will be published separately for some electrophoretic data on individual colloid particles and for some purely empirical trials on more general data.

This "analysis of reciprocals" by (28) is invariant under changes of scale of the observations, but not under more general linear transformations, whereas the analysis of variance is thus invariant. It should also be noted that the obvious parallel with the algebraic identity for the main effects and interactions in the analysis of variance for crossed classifications does not give independent components. An interaction term, such as

$$\sum_{i=1}^{N}\sum_{j=1}^{n}\left(\frac{1}{x_{ij}}-\frac{1}{x_{i.}}-\frac{1}{x_{.j}}+\frac{1}{x_{..}}\right)$$

in a commonly used notation, does not have a distribution of the chi-square type, since it has a finite probability of taking a negative value, and therefore this analogue of the analysis of variance is restricted to nested classifications.

6. Distribution of the reciprocal of an Inverse Gaussian variate. For some purposes it is convenient to work with the reciprocal of the Inverse Gaussian variate x, which will be denoted by y. For example, the analysis discussed in Section 5 can be expressed simply in terms of this variable. The weighted arithmetic means $x_{i.}$, $x_{..}$ of the values of x_{ij} are replaced by their reciprocals, which are the weighted harmonic means $\tilde{y}_{i.}$, $\tilde{y}_{..}$ of the values of y_{ij}. The analysis of the values of y_{ij}, corresponding to the algebraic identity (24) or (26), with $\psi(y) = y$ and M the harmonic mean, thus becomes

$$\begin{aligned}&\textstyle\sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij}-\tilde{y}_{..})\\ &\qquad=\textstyle\sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij}-\tilde{y}_{i.})+\sum_{i=1}^{N}n_i(\tilde{y}_{i.}-\tilde{y}_{..}).\end{aligned} \tag{29}$$

These sums of course have the same chi-square types of distribution as the expressions in terms of x_{ij}, $x_{i.}$, and $x_{..}$, to which they are equal. However, this analysis (29) is sufficiently easy to compute to be considered as a further practical analogue of the analysis of variance for certain purposes.

Some of the useful properties of the variate y follow in an obvious way from those of x, hardly justifying giving any special consideration to the family of distributions of y. However, the latter has some interesting features and a short account is therefore in order. Some of the results will be expressed in terms of x, since that variate is the primary object of this study.

The probability density function of y may be written

$$\exp\left\{-\frac{\lambda y}{2}+\frac{\lambda}{\mu}-\frac{\lambda}{2\mu^2 y}\right\}[\lambda/2\pi y]^{1/2}, \qquad 0<y<\infty \tag{30}$$

$$= \mu y f_2(y;\mu^{-1},\lambda\mu^{-2}) = \mu^2 y f_3(\mu y;1,\phi) = \mu^2 y f_4(\mu y;\phi,\phi). \tag{31}$$

The mode is at

$$y_{\text{mode}} = \frac{1}{2\lambda}[-1+(1+4\phi^2)^{1/2}] = \frac{1}{\mu}\left\{\left(1+\frac{1}{4\phi^2}\right)^{1/2}-\frac{1}{2\phi}\right\}.$$

The probability density at the mode is

$$\mu\{[1+(1+4\phi^2)^{1/2}]/4\pi\}^{1/2}\exp\{\phi-(\phi^2+\tfrac{1}{4})^{1/2}\},$$

which approaches $\mu/(2\pi e)^{1/2} = 0.241971\mu$ as its limit when ϕ or λ decreases to zero with a fixed value of μ.

Fig. 3 shows some examples of (30) plotted for $\lambda = 0, \frac{1}{4}, 1, 4, 16, 32$, with $\mu = 1$, for $0 < y < 3$, corresponding to $\frac{1}{3} < x < \infty$. The difference between Figs. 1 and 3 for small values of λ is rather striking.

Fig. 4 shows density curves from (30) in a form comparable with Fig. 2, having $\lambda = 1$ with $\mu = \frac{1}{4}, 1, 4$ for $0 < y < 7$. Thus the harmonic sample mean is a sufficient statistic for discriminating between the distributions of the family to which the curves in Fig. 4 belong, while the arithmetic mean is the corresponding statistic for Fig. 2 (cf. Section 3). Some consequences of using the arithmetic mean of observed values of y to estimate $1/\mu$ instead of using the harmonic mean are discussed in Section 7.

The moments about zero of y, which are the moments about zero of negative order for the Inverse Gaussian variate x, may be found by direct integration, using (3), or from other results. They may be found from those of x of positive order by using the relationship

$$E[(x/\mu)^{-r}] = E[(x/\mu)^{r+1}],$$

from which, in a notation applying to the variate x,

$$\mu'_{-r} = \mu'_{r+1}/\mu^{2r+1}. \tag{32}$$

Thus the moments of all positive and negative orders exist for an Inverse Gaussian variate (and for its reciprocal), in contrast to the situation with some other

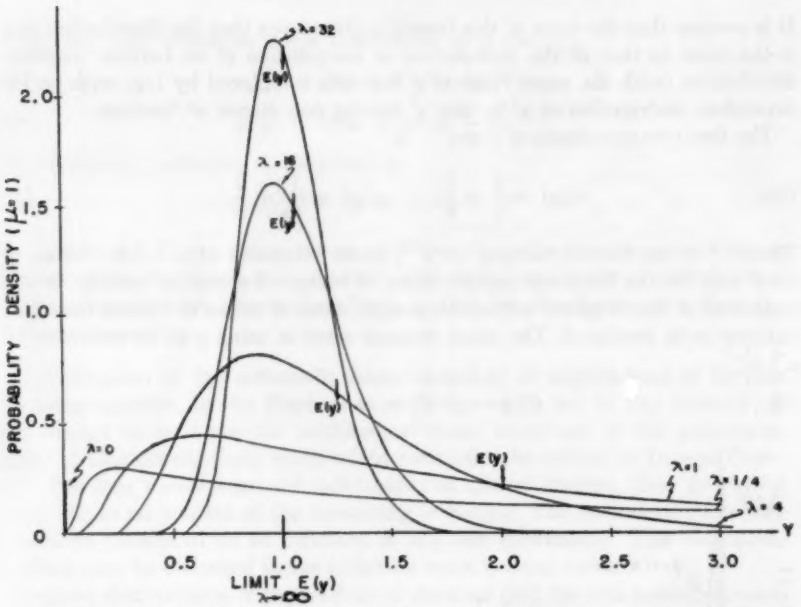


FIG. 3. Probability density curves for the reciprocal of an Inverse Gaussian variate with $\mu = 1$ for 6 values of λ or ϕ.

superficially similar distributions, such as the chi-square type. For reference, we have

$$\begin{aligned}
\mu'_{-1} &= \mu^{-1} + \lambda^{-1}, \\
\mu'_{-2} &= \mu^{-2} + 3\mu^{-1}\lambda^{-1} + 3\lambda^{-2}, \\
\mu'_{-3} &= \mu^{-3} + 6\mu^{-2}\lambda^{-1} + 15\mu^{-1}\lambda^{-2} + 15\lambda^{-3}, \\
\mu'_{-4} &= \mu^{-4} + 10\mu^{-3}\lambda^{-1} + 45\mu^{-2}\lambda^{-2} + 105\mu^{-1}\lambda^{-3} + 105\lambda^{-4}, \qquad (33) \\
\mu'_{-r} &= (2\lambda)^{-r} \sum_{s=0}^{r} \frac{(2r-s)!}{s!(r-s)!} (2\phi)^s \\
&= \mu^{-r} \sum_{s=0}^{r} \frac{(r+s)!}{s!(r-s)!} (2\phi)^{-s}.
\end{aligned}$$

The family (30) is of the Laplacian form as regards the variate y. Thus its cgf. can be found by a process of substituting alternative values for the parameters in the integral of (30), in a similar way to that used for deriving the corresponding function (8) for x. The result is that the logarithm of the Laplace transform of the density function of y is

$$L_y(t; \phi, \lambda) = \phi(1 - (1 + 2t\lambda^{-1})^{1/2}) - \tfrac{1}{2}\ln(1 + 2t\lambda^{-1}). \qquad (34)$$

It is curious that the form of this function (34) shows that the distribution of y is the same as that of the convolution or composition of an Inverse Gaussian distribution (with the same value of ϕ but with μ replaced by $1/\mu$) with an independent distribution of χ^2/λ, this χ^2 having one degree of freedom.

The first two cumulants of y are

$$\kappa_1(y) = \frac{1}{\mu} + \frac{1}{\lambda}, \qquad \kappa_2(y) = \frac{1}{\lambda\mu} + \frac{2}{\lambda^2}. \tag{35}$$

Thus λ^{-1} is the bias in using y (or x^{-1}) as an estimator of μ^{-1}. The variate y itself may be the harmonic sample mean of values of a similar variate, or the reciprocal of the weighted arithmetic sample mean of values of Inverse Gaussian variates as in Section 3. The mean squared error in using y as an estimator of μ^{-1} is

$$E[(y - \mu^{-1})^2] = (\phi + 3)\lambda^{-2}. \tag{36}$$

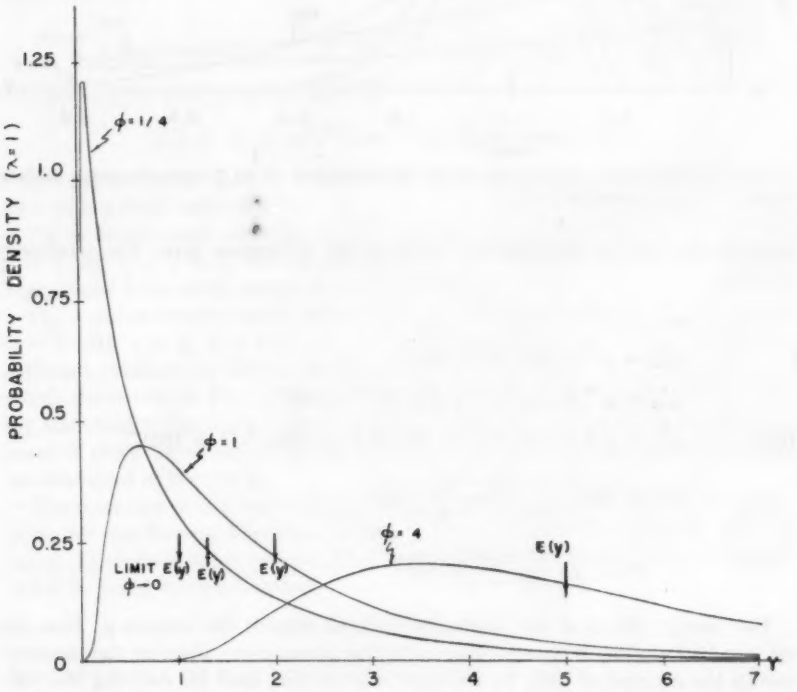


FIG. 4. Probability density curves for the reciprocal of an Inverse Gaussian variate with $\lambda = 1$ for 3 values of μ or ϕ.

The first two Fisherian shape coefficients of y are

$$\gamma_1(y) = (3\phi + 8)(\phi + 2)^{-3/2}, \tag{37}$$

$$\gamma_2(y) = 3(5\phi + 16)(\phi + 2)^{-2}. \tag{38}$$

The fractional coefficient of variation is

$$\gamma_{-1}(y)^{-1} = (\phi + 1)^{-1}(\phi + 2)^{1/2}. \tag{39}$$

The values of γ_{-1}^{-1}, γ_1, and γ_2 are smaller than the values of the corresponding characteristics of x. When ϕ is increased, both these and the shape coefficients of higher order approach zero as their limit, so that the distribution of y then approaches normality.

7. Estimation of the arithmetic mean reciprocal of expectations of Inverse Gaussian variates. In the physical experiments which led to this research, it was desired to estimate the arithmetical mean reciprocal of the population means of four distributions which might reasonably be treated as Inverse Gaussian. The four means were not individually of special interest, their inequality being due to an artifact of the measuring technique. The secondary parameter λ could be considered to be constant in any one experiment. This estimation problem may be discussed in the following more general terms:

Suppose that we have N populations of the kind (30), the i-th having parameter values μ_i, $\lambda_i = \lambda_0 w_i$, and one observation y_i from each population. Write $\bar{y} = \sum_{i=1}^{N} (w_i y_i)/\sum_{i=1}^{N} (w_i)$. It follows from the form of (34) that the distribution of $\bar{y}$ is the same as that of the simple arithmetic mean of N values of y taken from one distribution whose parameter values are

$$\mu^* = \sum_{i=1}^{N} (w_i)/\sum_{i=1}^{N} (w_i/\mu_i), \qquad \lambda^* = \lambda_0 \sum_{i=1}^{N} (w_i)/N.$$

We may also write $\phi^* = \lambda^*/\mu^*$. The distribution of $\bar{y}$ is the convolution of an Inverse Gaussian distribution, whose parameters in the form (1b) have the values $1/\mu^*$, $\lambda^* N$, with an independent distribution of $\chi^2/\lambda^* N$, this chi-square having N degrees of freedom. Because this belongs to a different Laplacian family it will not be studied in detail here. The results needed at present can be obtained from those already found.

From (35),

$$E(\bar{y}) = (1 + 1/\phi^*)/\mu^*, \tag{40}$$

and

$$E[(\bar{y} - 1/\mu^*)^2] = (\phi^* + 2 + N)/\lambda^{*2} N. \tag{41}$$

If the N values of μ_i were all equal, the harmonic sample mean

$$\bar{y} = \sum_{i=1}^{N} (w_i)/\sum_{i=1}^{N} (w_i/y_i)$$

would be a more precise estimator of the common value. The formula (36) would then give

$$E[(\bar{y} - 1/\mu)^2] = (\phi^* + 3/N)/\lambda^{*2}N, \tag{42}$$

which is less than (41) except in the trivial case of $N = 1$. The efficiency of $\bar{y}$, in these circumstances, can be measured by the ratio of the mean squared errors (42) and (41), or by the ratio, to N, of the modified value for N which needs to be substituted (without changing λ^* and ϕ^*) in (42) to give a mean squared error for $\bar{y}$ equal to (41). The former measure of efficiency is easier to calculate, but is slightly less than the latter. However, the difference is less than one per cent if $\phi^* > 18.7$.

Reverting to the estimation of μ^* when the values of μ_i may be unequal, we may attempt to improve the estimator $\bar{y}$ by adjusting it for the bias. When the N values of y_i are the harmonic sample means of the reciprocals of Inverse Gaussian variates, or the reciprocals of arithmetic sample means of Inverse Gaussian variates, separate estimates of the values of λ_i can be obtained from the variation exhibited within the samples, by using the appropriate form of (27). If the i-th sample contains n_i observations,

$$\hat{\lambda}_i^{-1} = \chi^2_{(n_i - 1 \text{ d.f.})}/\lambda_0 n_i w_i ,$$

this distribution being independent of y_i. On weighting these estimators suitably and on writing $\sum_{i=1}^{N} (n_i - 1) = D$, we have, as an unbiased estimator of λ_0^{-1} of minimum variance,

$$\sum_{i=1}^{N} (\hat{\lambda}_i^{-1} n_i w_i)/D = \chi^2_{(D \text{ d.f.})}/\lambda_0 D.$$

Thus, for an unbiased estimator of $1/\mu^*$, we get

$$y' = \sum_{i=1}^{N} w_i(y_i - n_i N/\hat{\lambda}_i D)/\sum_{i=1}^{N} (w_i). \tag{43}$$

Then

$$E[(y' - 1/\mu^*)^2] = (\phi^* + 2 + 2N/D)/\lambda^{*2}N. \tag{44}$$

This mean squared error will always be less than that (41) of $\bar{y}$ if $D > 2$, which will be true of most experiments. However, unless ϕ^* is close to or less than unity, which seems unlikely to occur, the statistical superiority of y' over $\bar{y}$ is of very minor importance.

8. Acknowledgments. Much of this work was done under a Senior Scholarship at the University of Reading, England, followed by financial assistance from the British Empire Cancer Campaign. The recent developments were made under the sponsorship of the National Science Foundation. The figures were drawn by Mrs. D. Hamilton from tables of values which were computed by Mr. G. Zorbalas.

REFERENCES

[1] M. C. K. Tweedie, "Inverse statistical variates," *Nature* (London), Vol. 155 (1945), p. 453.

[2] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, 1944.

[3] E. Schrödinger, "Zur Theorie der Fall- und Steigversuche an Teilchen mit Brownscher Bewegung", *Physikalische Zeitschrift*, Vol. 16 (1915), pp. 289–295.

[4] A. Wald, "On cumulative sums of random variables", *Ann. Math. Stat.*, Vol. 15 (1944), pp. 283–296.

[5] M. C. K. Tweedie, "A mathematical investigation of some electrophoretic measurements on colloids," unpublished thesis for M.Sc. degree, University of Reading, England, 1941.

[6] M. C. K. Tweedie, "Functions of a statistical variate with given means, with special reference to Laplacian distributions," *Proc. Cambridge Philos. Soc.*, Vol. 43 (1947), pp. 41–49.

[7] E. J. Scott, *Transform Calculus with an Introduction to Complex Variables*, Harper, New York, 1955.

[8] D. V. Widder, *The Laplace Transform*, Princeton University Press, 1941.

[9] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, 1950.

VARIANCES OF VARIANCE COMPONENTS: III. THIRD MOMENTS IN A BALANCED SINGLE CLASSIFICATION[1]

By John W. Tukey

Princeton University

1. Summary. The methods used in earlier papers of this series [2], [3] are extended from variances to third moments, and applied to the third moment about the mean (= third cumulant) of the usual estimate of the between variance component in a balanced single classification. The result is moderately complex, but manageable.

2. Introduction. The technique of this paper grows directly out of those used for variances in the earlier papers in this series [2], [3]. We assume familiarity with the terminology and notations used there.

We begin by discussing the third moment of the variance of a sample as an illustration of problems and technique, and then pass on directly to the main problem. We need the multiplication formulas (see Wishart [4])

$$k_2^2 = k_{22} + \frac{1}{n} k_4 + \frac{2}{n-1} k_{22},$$

$$k_2^3 = k_{222} + \frac{3(n+3)}{n(n-1)} k_{24} + \frac{6n+2}{(n-1)^2} k_{222} + \frac{1}{n^2} k_6 + \frac{4(n-2)}{n(n-1)^2} k_{33}.$$

$$(k_2')^3 = k_{222}' + 3\frac{N+3}{N(N-1)} k_{24}' + \frac{6N+2}{(N-1)^2} k_{222}' + \frac{1}{N^2} k_6' + 4\frac{(N-2)}{(N-1)^2} k_{33}',$$

$$k_2' k_{22}' = k_{222}' + \frac{2}{N} k_{24}' + \frac{4}{N-1} k_{222}' - \frac{2}{N(N-1)} k_{33}',$$

$$k_2' k_4' = k_{24}' + \frac{1}{N} k_6' + \frac{8}{N-1} k_{24}' + \frac{6}{N-1} k_{33}'$$

3. The variance. We can now proceed to write down the third moment about the mean (that is the *third cumulant*) of the variance of a sample drawn from a finite population. There are various ways to do this, but we shall begin with one resembling the first method we used [1] to get the variance of the variances (we refer to using primes for population quantities for three sections):

$$\begin{aligned}\text{ave}\,\{k_2 - k_2'\}^3 &= \text{ave}\,\{k_2^3 - 3k_2^2 k_2' + 3k_2(k_2')^2 - (k_2')^3\} \\ &= \text{ave}\,\{k_2^3\} - 3k_2'\,\text{ave}\,\{k_2^2\} + 2(k_2')^3\end{aligned}$$

Received April 6, 1955.

[1] Prepared in connection with research sponsored by the Office of Naval Research and based on part of Memorandum Report 45, "Finite Sampling Simplified," Statistical Research Group, Princeton University, which was written while the author was a Fellow of the John Simon Guggenheim Foundation.

$$= \text{ave}\left\{k_{222} + \frac{3(n+3)}{n(n-1)}\,k_{24} + \frac{6n+2}{(n-1)^2}\,k_{222} + \frac{1}{n^2}\,k_6 + \frac{4(n-2)}{n(n-1)^2}\,k_{33}\right\}$$

$$- 3k_2' \,\text{ave}\left\{k_{22} + \frac{1}{n}\,k_4 + \frac{2}{n-1}\,k_{22}\right\} + 2k_2'^3$$

$$= k_{222}' + \frac{3(n+3)}{n(n-1)}\,k_{24}' + \frac{6n+2}{(n-1)^2}\,k_{222}' + \frac{1}{n^2}\,k_6' + \frac{4(n-2)}{n(n-1)^2}\,k_{33}'$$

$$- 3k_2'k_{22}' - \frac{3}{n}\,k_2'k_4' - \frac{6}{n-1}\,k_2'k_{22}' + 2(k_2')^3$$

$$= k_{222}' + \frac{3(n+3)}{n(n-1)}\,k_{24}' + \frac{6n+2}{(n-1)^2}\,k_{222}' + \frac{1}{n^2}\,k_6' + \frac{4(n-2)}{n(n-1)^2}\,k_{33}'$$

$$- 3k_{222}' - \frac{6}{N}\,k_{24}' - \frac{12}{N-1}\,k_{222}' + \frac{6}{N(N-1)}\,k_{33}' - \frac{3}{n}\,k_{24}'$$

$$- \frac{3}{nN}\,k_6' - \frac{24}{(N-1)n}\,k_{24}' - \frac{18}{(N-1)n}\,k_{33}' - \frac{6}{n-1}\,k_{222}'$$

$$- \frac{12}{(n-1)N}\,k_{24}'$$

$$- \frac{24}{(n-1)(N-1)}\,k_{222}' + \frac{12}{(n-1)N(N-1)}\,k_{33}' + 2k_{222}'$$

$$+ 6\,\frac{N+3}{N(N-1)}\,k_{24}' + 2\,\frac{6N+2}{(N-1)^2}\,k_{222}' + \frac{2}{N^2}\,k_6' + \frac{8(N-2)}{N(N-1)^2}\,k_{33}'.$$

This may be written as

$$\left\{\frac{1}{n^2} - \frac{3}{nN} + \frac{2}{N^2}\right\}k_6' + \left\{\frac{4}{n(n-1)} - \frac{18}{(n-1)N} + \frac{14}{N(N-1)}\right\}k_{33}'$$

$$- \left\{\frac{4}{n(n-1)^2} - \frac{12}{(n-1)N(N-1)} + \frac{8}{N(N-1)^2}\right\}k_{33}'$$

$$+ \left\{\frac{8}{(n-1)^2} - \frac{24}{(n-1)(N-1)} + \frac{16}{(N-1)^2}\right\}k_{222}'$$

$$+ \left\{\frac{12}{n(n-1)} - \frac{12}{(n-1)N} - \frac{24}{(N-1)n} + \frac{24}{N(N-1)}\right\}k_{24}',$$

which vanishes for $n = N$, just as it should.

Now it is even clearer for the third cumulant (which might perhaps be called the *skewmulant*) than it was for the variance, that direct calculation would be long, tedious, boring, and, because of its length, likely to be erroneous for any component in an analysis of variance situation. We must go to more ingenious methods.

4. Structure. We notice that the third cumulant of any algebraic function of a random sample takes the form

$$\text{ave}\,\{u^4\} - 3\,\text{ave}\,\{u\}\,\text{ave}\,\{u^2\} + 2(\text{ave}\,\{u\})^3 = \alpha(n) - 3\delta\beta(n)\gamma(n) + 2\delta^4,$$

where $\alpha(n)$, $\beta(n)$, and $\gamma(n)$ depend on n alone, and δ is independent of n and N, all four being expressable linearly in polykays. The only appearance of N comes from the multiplication formulas involved in

$$\beta(n)\cdot\gamma(n)$$

and

$$\delta\cdot\delta\cdot\delta.$$

The final result has the form

$$A(n) + \sum B(n)C(N) + D(N),$$

where the functions of N come from

$$-3\,\text{ave}\,\{u\}\,\text{ave}\,\{u^2\} + 2\{\text{ave}\,\{u\}\}^3,$$

which we may suppose already known.

We can again proceed by finding these two terms first, and then using special populations to determine the coefficients in $A(n)$.

5. The variance-second method. We have

$$\begin{aligned}
\text{ave}\,\{k_2\} &= k_2', \\
\text{ave}\,\{k_2^2\} &= (\text{ave}\,\{k_2\})^2 + \text{var}\,\{k_2\} \\
&= (k_2')^2 + \left(\frac{1}{n} - \frac{1}{N}\right)k_4' + \left(\frac{2}{n-1} - \frac{2}{N-1}\right)k_{22}' \\
&= \left\{1 + \frac{1}{n-1}\right\}k_{22}' + \frac{1}{n}\,k_4',
\end{aligned}$$

then $\sum B(n)C(N) + D(N)$ must come from

$$-\frac{6}{n-1}\,k_{22}'k_2' - \frac{3}{n}\,k_4'k_2' + 2(k_2')^3 - 3k_{22}'k_2',$$

whence

$$\sum B(n)C(N) = -\frac{6}{n-1}\left\{\frac{2}{N}\,k_{24}' + \frac{4}{N-1}\,k_{222}' - \frac{2}{N(N-1)}\,k_{33}'\right\} - \frac{3}{n}\left\{\frac{1}{N}\,k_6' + \frac{8}{N-1}\,k_{24}' + \frac{6}{N-1}\,k_{33}'\right\},$$

$$D(N) = 2\left\{3\,\frac{N+3}{N(N-1)}\,k_{24}' + \frac{6N+2}{(N-1)^2}\,k_{222}' + \frac{1}{N^2}\,k_6' + 4\,\frac{N-2}{N(N-1)^2}\,k_{33}'\right\} - 3\left\{\frac{2}{N}\,k_{24}' + \frac{4}{N-1}\,k_{222}' - \frac{2}{N(N-1)}\,k_{33}'\right\},$$

thus, the part of the third cumulant of the usual estimate of k_2 which depends on the population size, N, is

$$\left\{\frac{2}{N^2} - \frac{3}{nN}\right\} k_6'$$
$$+\left\{\frac{14}{N(N-1)} - \frac{18}{n(N+1)} - \frac{8}{N(N-1)^2} + \frac{12}{(n-1)N(N-1)}\right\} k_{33}'$$
$$\left\{\frac{24}{N(N-1)} - \frac{12}{N(n-1)} - \frac{24}{n(N-1)}\right\} k_{24}'$$
$$+\left\{\frac{16}{(N-1)^2} - \frac{24}{(n-1)(N-1)}\right\} k_{222}'.$$

When $n = N$ this reduces to the negative of

$$(*) \quad \frac{1}{n^2} k_6' + \left\{\frac{4}{n(n-1)} - \frac{4}{n(n-1)^2}\right\} k_{33}' + \frac{12}{n(n-1)} k_{24}' + \frac{8}{(n-1)^2} k_{222}',$$

and since the k_2 estimate is constant when $n = N$ the additional terms not involving N must be those just given in (*).

6. The between component in a balanced single classification. We now come back to the balanced single classification and drop the primes. Hence the column contribution is drawn from n $k_1, k_{11}, \cdots$, and the error contribution is drawn from $N, K_1, K_{11}, \cdots$. We are going to find the third cumulant of the between component in sampling from an arbitrary finite population.

When we express this third cumulant multilinearly in the k's and K's, there may be terms involving

(1) $k_6, k_{24}, k_{33}, k_{222}$,

(2) $k_4K_2, k_{22}K_2, k_3K_3, K_2K_4, k_2K_{22}$,

(3) $K_6, K_{24}, K_{33}, K_{222}$,

and no others (because of homogeneity and invariance under translation). The discussion just given for the variance of a sample applies with minor changes to the case where the errors are constant. This determines the coefficient of the terms involving the monomials in (1) to be the same as for the third cumulant of the variance, with c replacing n.

We go next to the terms involving the monomials in (3). Suppose first that the column contributions are constant. If we take a minimal unit population for the individual contributions, the between component is constant. Thus, the desired third cumulant, and everything it could involve except K_6, vanishes. Thus the coefficient of K_6 is itself zero.

In order to deal with the next coefficients, we can use a population with two non-zero values for the individual contributions. But if both values are alike, we cannot distinguish between the coefficients of K_{33} and K_{24}. So we use a population of rc values $1, t, 0, \cdots, 0$, for which

$$K_{33} = \frac{2t^3}{rc(rc-1)}, \qquad K_{24} = \frac{t^2 + t^4}{rc(rc-1)}.$$

There are two cases to consider

CASE 1. (Probability $(r-1)/(rc-1)$.) One $x = 1$ and another $x = t$ in the same column; others, zero.

CASE 2. (Probability $r(c-1)/(rc-1)$.) One $x = 1$ and another $x = t$ in different columns; others, zero.

The corresponding analyses of variance are

CASE 1

	DF	SS	MS	CMS
Between	$c-1$	$\dfrac{(t+1)^2(c-1)}{rc}$	$\dfrac{(t+1)^2}{rc}$	$\dfrac{2t}{rc(r-1)}$
Within	$c(r-1)$	$t^2+1-\dfrac{(t+1)^2}{r}$	$\dfrac{t^2+1}{rc}-\dfrac{2t}{rc(r-1)}$	$\dfrac{t^2+1}{rc}-\dfrac{2t}{rc(r-1)}$

CASE 2

	DF	SS	MS	CMS
Between	$c-1$	$\dfrac{(c-1)(t^2+1)}{rc}-\dfrac{2t}{rc}$	$\dfrac{t^2+1}{rc}-\dfrac{2t}{rc(c-1)}$	$\dfrac{-2t}{r^2c(c-1)}$
Within	$c(r-1)$	$\dfrac{(r-1)(t^2+1)}{r}$	$\dfrac{t^2+1}{rc}$	$\dfrac{t^2+1}{rc}$

Thus the mean of the between component vanishes, as it should, and the third cumulant reduces to the third moment about zero, which is

$$\frac{r-1}{rc-1}\left\{\frac{2}{rc}\frac{t}{r-1}\right\}^3+\frac{r(c-1)}{rc-1}\left\{-\frac{2}{rc}\frac{t}{r(c-1)}\right\}^3$$

$$=\frac{8t^3}{r^3c^3(rc-1)}\left\{\frac{1}{(r-1)^2}-\frac{1}{r^2(c-1)^2}\right\}.$$

This has t^3 as a factor, and does not involve t^2+t^4, hence the coefficient of K_{24} is zero, while that of K_{33} is

$$\frac{4}{r^2c^2}\left\{\frac{1}{(r-1)^2}-\frac{1}{r^2(c-1)^2}\right\},$$

which, as we might have suspected, is small when the two sets of degrees of freedom are of nearly the same size.

We can deal with the coefficient of K_{222}, and certain of the other terms, by resorting to normal theory. In this case, the two mean squares become independent and their third cumulants are

$$8(K_2+rk_2)^3/(c-1)^2, \qquad 8K_2^3/c^3(r-1)^2,$$

so that the third cumulant of the between component is

$$\frac{8}{r^3}\left\{\frac{(K_2 + rk_2)^2}{(c-1)^2} - \frac{K_2^2}{c^2(r-1)^2}\right\} = \frac{8}{r^3}\left(\frac{1}{(c-1)^2} - \frac{1}{c^2(r-1)^2}\right)K_2^2 + \frac{3r}{(c-1)^2}K_2^2 k_2 + \frac{3r^2}{(c-1)^2}K_2 k_2^2 + \frac{r^3}{(c-1)^2}k_2^3\Bigg\}$$

$$= \frac{8}{r^3}\left(\frac{1}{(c-1)^2} - \frac{1}{c^2(r-1)^2}\right)K_{222} + \frac{24}{r^2(c-1)^2}K_{22}k_2 + \frac{24}{r(c-1)^2}K_2 k_{22} + \frac{8}{(c-1)}k_{222}.\Bigg\}$$

Thus, we have determined the coefficients of K_{222}, $K_{22}k_2$, and K_2k_{22}, and have confirmed the part of the coefficient of k_{222} which is independent of N.

There remain the coefficients of k_4K_2, k_3K_3, and k_2K_4. These we shall seek by taking a minimal unit population for the column contributions and a minimal population with one nonzero value equal to t for the error. We have

$$k_4 = 1/c, \quad k_3 = 1/c, \quad k_2 = 1/c,$$
$$K_4 = t^4/rc, \quad K_3 = t^3/rc, \quad K_2 = t^2/rc,$$

so that

$$k_4K_2 = \frac{t^2}{rc^2}, \quad k_3K_3 = \frac{t^3}{rc^2}, \quad k_2K_4 = \frac{t^4}{rc^2}.$$

We have the usual two cases to deal with.

CASE 3. (Probability $1/c$). One $x = t + 1$, others in column = 1, others zero.

CASE 4. (Probability $(c-1)/c$). One column of x's = 1, one other $x = t$, all others zero.

Corresponding analyses of variance are

CASE 3

	DF	SS	MS	CMS
Between	$c-1$	$\frac{(t+r)^2(c-1)}{rc}$	$\frac{(t+r)^2}{rc}$	$\frac{1}{c} + \frac{2t}{rc}$
Within	$c(r-1)$	$t^2\frac{r-1}{r}$	$\frac{t^2}{rc}$	$\frac{t^2}{rc}$

CASE 4

	DF	SS	MS	CMS
Between	$c-1$	$\left(r + \frac{t^2}{r}\right)\frac{c-1}{c} - \frac{2t}{c}$	$\frac{t^2+r^2}{rc} - \frac{2t}{c(c-1)}$	$\frac{1}{c} - \frac{2t}{rc(c-1)}$
Within	$c(r-1)$	$t^2\left(\frac{r-1}{r}\right)$	$\frac{t^2}{rc}$	$\frac{t^2}{rc}$

The mean between component is $1/c$, as it should be, and its third moment about the mean is

$$\frac{1}{c}\left(\frac{2t}{rc}\right)^3 + \frac{c-1}{c}\left(-\frac{2t}{rc(c-1)}\right)^3 = \frac{8t^3}{r^3c^3}\left\{1 - \frac{1}{(c-1)^2}\right\}.$$

The terms in t^2 and t^4 are conspicuous in their absence, and hence the coefficients of k_4K_2 and k_2K_4 vanish. The coefficient of k_3K_3 is

$$\frac{4}{r^2c^2}\left\{1 - \frac{1}{(c-1)^2}\right\}.$$

We can now reassemble all our coefficients and write down the third cumulant of the between component in a balanced single classification with c categories and r units per category. The result is

$$\begin{aligned}
&\left\{\frac{1}{c^2} - \frac{3}{cn} + \frac{2}{n^2}\right\} k_6 + \left\{\frac{4}{c(c-1)} - \frac{18}{(c-1)n} + \frac{14}{n(n-1)}\right\} k_{33} \\
&\quad - \left\{\frac{4}{c(c-1)^2} - \frac{12}{(c-1)n(n-1)} + \frac{8}{n(n-1)^2}\right\} k_{33} \\
&\quad + \left\{\frac{12}{c(c-1)} - \frac{12}{(c-1)n} - \frac{24}{c(n-1)} + \frac{24}{n(n-1)}\right\} k_{24} \\
&\quad + \left\{\frac{8}{(c-1)^2} - \frac{24}{(c-1)(n-1)} + \frac{16}{(n-1)^2}\right\} k_{222} \\
&\quad + \frac{24}{r(c-1)^2} k_{22}K_2 + \frac{4}{r^2c^2}\left\{1 - \frac{1}{(c-1)^2}\right\} k_3K_3 + \frac{24}{r^2(c-1)^2} k_2K_{22} \\
&\quad + \frac{4}{r^2c^2}\left\{\frac{1}{(r-1)^2} - \frac{1}{r^2(c-1)^2}\right\} K_{33} + \frac{8}{r^3}\left\{\frac{1}{(c-1)^2} - \frac{1}{c^2(r-1)^2}\right\} K_{222},
\end{aligned}$$

for populations of n column contributions and N error contributions.

For infinite populations this reduces slightly, and becomes

$$\begin{aligned}
&\frac{1}{c^2} k_6 + \frac{4(c-2)}{c(c-1)^2} k_{33} + \frac{12}{c(c-1)} k_{24} + \frac{8}{(c-1)^2} k_{222} + \frac{24}{r(c-1)^2} k_{22}K_2 \\
&\quad + \frac{4}{r^2c^2}\left\{1 - \frac{1}{(c-1)^2}\right\} k_3K_3 + \frac{24}{r^2(c-1)^3} k_2K_{22} \\
&\quad + \frac{4}{r^2c^2}\left\{\frac{1}{(r-1)^2} - \frac{1}{r^2(c-1)^2}\right\} K_{33} + \frac{8}{r^3}\left\{\frac{1}{(c-1)^2} - \frac{1}{c^2(r-1)^2}\right\} K_{222}.
\end{aligned}$$

REFERENCES

[1] JOHN W. TUKEY, "Some sampling simplified," *J. Amer. Statist. Assn.*, Vol. 45 (1950), 501–519.

[2] JOHN W. TUKEY, "Variances of variance components: I. Balanced designs," *Ann. Math. Stat.*, Vol. 27 (1956), 722–736.

[3] JOHN W. TUKEY, "Variances of variance components: II. Unbalanced single classifications." *Ann. Math. Stat*, Vol. 28 (1957), 43–56.

[4] JOHN WISHART, "Moment coefficients of the k-statistics in samples from a finite population," *Biometrika*, Vol. 39, pp. 1–13.

SOME REMARKS ON SYSTEMATIC SAMPLING[1]

By Werner Gautschi[2]

University of California, Berkeley

1. Introduction and summary. Consider a finite population consisting of N elements $y_1, y_2, \cdots, y_N$. Throughout the paper we will assume that $N = nk$. A systematic sample of n elements is drawn by choosing one element at random from the first k elements $y_1, \cdots, y_k$, and then selecting every kth element thereafter. Let $y_{ij} = y_{i+(j-1)k} (i = 1, \cdots, k; j = 1, \cdots, n)$; obviously systematic sampling is equivalent to selecting *one* of the k "clusters"

$$C_i = \{y_{ij}; j = 1, \cdots, n\}$$

at random. From this it follows that the sample mean $\bar{y}_i = 1/n \sum_{j=1}^{n} y_{ij}$ is an unbiased estimate for the population mean $\bar{y} = 1/N \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}$ and that $\operatorname{Var} \bar{y}_i = 1/k \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2$. We will denote this variance by $V_{sy}^{(1)}$ indicating by the superscript that only one cluster is selected at random. $V_{sy}^{(1)}$ can be written as

$$V_{sy}^{(1)} = S^2 - \frac{1}{k} \sum_{i=1}^{k} S_i^2, \quad \text{where} \quad S^2 = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y})^2,$$

(1)

$$S_i^2 = \frac{1}{n} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2.$$

It is natural to compare systematic sampling with stratified random sampling, where *one* element is chosen independently in each of the n strata $\{y_1, \cdots, y_k\}$, $\{y_{k+1}, \cdots, y_{2k}\}, \cdots$, and with simple random sampling using sample size n. The corresponding variances of the sample mean will be denoted by $V_{st}^{(1)}$ $V_{ran}^{(n)}$ respectively.

We consider now the following generalization of systematic sampling which appears to have been suggested by J. Tukey (see [3], p. 96, [4], [5]). Instead of choosing at first only one element at random we select a simple random sample of size s (without replacement) from the first k elements and then every kth element following those selected. In this way we obtain a sample of ns elements and, if $i_1, i_2, \cdots, i_s$ are the serial numbers of the elements first chosen, the sample mean $1/s(\bar{y}_{i_1} + \cdots + \bar{y}_{i_s})$ can be used as an estimate for the population mean. This sampling procedure is clearly equivalent to drawing a simple random sample of size s from the k clusters $C_i (i = 1, \cdots, k)$. It therefore follows (see, for example, [2], Chapter 2.3 to 2.4) that the sample mean is an unbiased estimate for the population mean and that its variance, which we denote by $V_{sy}^{(s)}$, is given by[3]

Received June 12, 1956.

[1] This investigation was supported (in part) by a research grant (RG-3666) from the Institutes of Health, Public Health Service.

[2] Now at Ohio State University

[3] This formula is not new, but appeared already in [6] and, more recently, in [5].

$$V_{sy}^{(s)} = \frac{k-s}{ks}\frac{1}{k-1}\sum_{i=1}^{k}(\bar{y}_i - \bar{y})^2 = \frac{1}{s}\frac{k-s}{k-1}V_{sy}^{(1)}. \tag{2}$$

Again, it is natural to compare this sampling procedure with stratified random sampling, where a simple random sample of size s is drawn independently in each of the n strata $\{y_1, \cdots, y_k\}$, $\{y_{k+1}, \cdots, y_{2k}\}$, $\cdots$ or with simple random sampling employing sample size ns. We denote the corresponding variances of the sample mean (which in both cases is an unbiased estimate for the population mean) by $V_{st}^{(s)}$, $V_{ran}^{(ns)}$ respectively. From well-known variance formulae (see, for example, [2], Chapters 2.4 and 5.3) it follows that

$$\begin{aligned} V_{st}^{(s)} &= \frac{1}{s}\frac{k-s}{k-1}V_{st}^{(1)}, \\ V_{ran}^{(ns)} &= \frac{N-ns}{s(N-n)}V_{ran}^{(n)} = \frac{1}{s}\frac{k-s}{k-1}V_{ran}^{(n)}. \end{aligned} \tag{3}$$

Thus the relative magnitudes of the three variances $V_{sy}^{(s)}$, $V_{st}^{(s)}$, $V_{ran}^{(ns)}$ are the same as for $V_{sy}^{(1)}$, $V_{st}^{(1)}$, $V_{ran}^{(n)}$, of which comparisons were made for several types of populations by W. G. Madow and L. H. Madow [6] and W. G. Cochran [1]. Some of the results will be reviewed in Section 3.

The object of this note is to compare systematic sampling with s random starts, as described above, with systematic sampling employing only one random start but using a sample of the same size ns. To make this comparison we obviously have to assume that k is an integral multiple of s, say $k = ls$. The latter procedure then consists in choosing one element at random from the first l elements $\{y_1, \cdots, y_l\}$ and selecting every lth consecutive element. We denote the variances of the sample mean of the two procedures by $V_k^{(s)}$, $V_l^{(1)}$ respectively, indicating by the subscript the size of the initial "counting interval." (In our notation $V_{sy}^{(s)} \equiv V_k^{(s)}$.) We shall show in Section 4 that $V_l^{(1)} = V_k^{(s)}$ in the case of a population "in random order," but $V_l^{(1)} < V_k^{(s)}$ for a population with a linear trend or with a positive correlation between the elements which is a decreasing convex function of their distance apart. Some numerical results on the relative precision of the two procedures will be given in Section 5 for the case of a large population with an exponential correlogram.

2. Acknowledgment. I wish to express my debt to Professor W. Kruskal for having brought the question treated in this note to my attention.

3. Cochran's approach. Extension of Cochran's results to systematic sampling with multiple random starts. Instead of considering a particular single population $\{y_1, y_2, \cdots, y_N\}$ we assume, following Cochran [1], [2], Chapter 8, that the y_i's are drawn from an infinite population having some specified properties. We are then interested in comparing the expected variance $E(V \mid y_1, \cdots, y_N)$ rather than $(V \mid y_1, \cdots, y_N)$ for the sampling procedures under consideration. More specifically, we consider the following three types of populations.

(i) *Population in random order.* The variates y_i are assumed to be uncor-

related and to have the same expectations. The variances may change with i

$$
\begin{aligned}
&Ey_i = \mu, \qquad E(y_i - \mu)^2 = \sigma_i^2 \qquad (i = 1, \cdots, N); \\
&E(y_i - \mu)(y_j - \mu) = 0 \qquad (i \neq j).
\end{aligned} \tag{4}
$$

It is not difficult to show ([2], Chapter 8.5) that in this case

$$
EV_{sy}^{(1)} = EV_{st}^{(1)} = EV_{ran}^{(n)} = \frac{N-n}{N}\frac{\sigma^2}{n} = \frac{k-1}{k}\frac{\sigma^2}{n}, \tag{5}
$$

where $\sigma^2 = \sum_{i=1}^{N} \sigma_i^2/N$.

(ii) *Population with a linear trend.* We assume that the y_i's are uncorrelated variates whose expectations change linearly in i, more precisely

$$
\begin{aligned}
&Ey_i = \alpha + \beta i, \qquad \operatorname{Var} y_i = \sigma^2 \qquad (i = 1, 2, \cdots, N), \\
&\operatorname{Cov}(y_i, y_j) = 0 \qquad (i \neq j).
\end{aligned} \tag{6}
$$

Applying standard linear regression theory (see, for instance, [7], Chapter 14.2) to the sum of squares in (1), it is easily found that

$$
EV_{sy}^{(1)} = \frac{N-n}{Nn}\sigma^2 + \beta^2\frac{k^2-1}{12} = \frac{k-1}{nk}\sigma^2 + \beta^2\frac{k^2-1}{12}. \tag{8}
$$

In a similar way we obtain

$$
\begin{aligned}
EV_{st}^{(1)} &= \frac{k-1}{nk}\sigma^2 + \beta^2\frac{k^2-1}{12n}, \\
EV_{ran}^{(n)} &= \frac{k-1}{nk}\sigma^2 + \beta^2\frac{(k-1)(nk+1)}{12}.
\end{aligned} \tag{9}
$$

Thus

$$
EV_{st}^{(1)} \leqq EV_{sy}^{(1)} \leqq EV_{ran}^{(n)}, \tag{10}
$$

with equality only if $n = 1$.

(iii) *Population with serial correlation.* It is assumed that two elements y_i, y_j are positively correlated with a correlation which depends only on the "distance" $z = |j - i|$ and which decreases as z increases. The mean and variance of all the y_i are supposed to be constant

$$
\begin{aligned}
&Ey_i = \mu, \qquad E(y_i - \mu)^2 = \sigma^2 \qquad (i = 1, 2, \cdots, N), \\
&E(y_i - \mu)(y_{i+z} - \mu) = \rho_z\sigma^2,
\end{aligned} \tag{11}
$$

where $\rho_{z_1} \geqq \rho_{z_2} \geqq 0$ for $z_1 < z_2$. For this type of population Cochran [1] obtained the following results relevant to our purpose:

$$
\begin{aligned}
EV_{sy}^{(1)} = \frac{k-1}{N}\sigma^2\Bigg\{1 &- \frac{2}{N(k-1)}\sum_{z=1}^{N-1}(N-z)\rho_z \\
&+ \frac{2k}{n(k-1)}\sum_{z=1}^{n-1}(n-z)\rho_{kz}\Bigg\},
\end{aligned} \tag{12}
$$

(13) $EV_{st}^{(1)} \leqq EV_{ran}^{(n)}$,

(14) $EV_{sy}^{(1)} \leqq EV_{st}^{(1)}$,

(14) applying if, in addition, ρ_s is convex downwards.

In virtue of (2) and (3) all the results (5), (10), (13) and (14) carry over immediately to the more general sampling procedure discussed in Section 1 and, moreover, the relative sizes of the variances $V_{sy}^{(s)}$, $V_{st}^{(s)}$, $V_{ran}^{(ns)}$ remain the same as those of $V_{sy}^{(1)}$, $V_{st}^{(1)}$, $V_{ran}^{(n)}$. Numerical results of the relative precision

$$EV_{st}^{(1)}/EV_{sy}^{(1)}$$

were given by Cochran [1] for populations with a linear and exponential correlogram.

4. Comparison of systematic sampling and systematic sampling with multiple random starts.

(i) *Population in random order.* From (5), replacing k by l and n by ns, we obtain

$$EV_l^{(1)} = \frac{l-1}{lns}\sigma^2 = \frac{l-1}{N}\sigma^2.$$

On the other hand, by (2) and (5), remembering that $k = sl$,

$$EV_k^{(s)} = \frac{1}{s}\frac{k-s}{k-1}\frac{k-1}{k}\frac{\sigma^2}{n} = \frac{l-1}{N}\sigma^2.$$

Thus

(15) $$EV_l^{(1)} = EV_k^{(s)}.$$

(ii) *Population with linear trend.* By (2) and (8)

$$EV_l^{(1)} = \frac{l-1}{N}\sigma^2 + \beta^2\frac{(l-1)(l+1)}{12},$$

$$EV_k^{(s)} = \frac{1}{s}\frac{k-s}{k-1}\left[\frac{k-1}{nk}\sigma^2 + \beta^2\frac{k^2-1}{12}\right] = \frac{l-1}{N}\sigma^2 + \beta^2\frac{(l-1)(ls+1)}{12}.$$

Hence

(16) $$EV_l^{(1)} \leqq EV_k^{(s)}$$

with equality only if $s = 1$.

Both these results are, of course, to be expected intuitively. The comparison of $V_l^{(1)}$ and $V_k^{(s)}$ is, perhaps, mostly relevant for a population with a convex decreasing correlogram, since in this case $EV_l^{(1)}$ turns out to be the smallest among all the variances $EV_l^{(1)}$, $EV_k^{(s)}$, $EV_{st}^{(ns)}$, $EV_{ran}^{(ns)}$.

(iii) *Population with serial correlation.* From (12) and (2),

$$EV_l^{(1)} = \frac{l-1}{N}\sigma^2\left\{1 - \left[\frac{2}{N(l-1)}\sum_{z=1}^{N-1}(N-z)\rho_z\right.\right.$$

$$- \frac{2l}{ns(l-1)} \sum_{z=1}^{ns-1} (ns - z)\rho_{lz} \Bigg\} \tag{17}$$

$$= \frac{l-1}{N} \sigma^2\{1 - L_1\},$$

$$EV_k^{(s)} = \frac{l-1}{N} \sigma^2 \Bigg\{1 - \Bigg[\frac{2}{N(k-1)} \sum_{z=1}^{N-1} (N - z)\rho_z$$

$$- \frac{2k}{n(k-1)} \sum_{z=1}^{n-1} (n - z)\rho_{kz} \Bigg] \Bigg\} \tag{18}$$

$$= \frac{l-1}{N} \sigma^2\{1 - L_2\}.$$

It is easy to check that both L_1 and L_2 are linear forms in the ρ_z's in each of which the sum of coefficients is equal to 1. Hence, in order to show that $EV_l^{(1)} \leqq EV_k^{(s)}$, it is enough to prove that

$$L = L_1 - L_2 \geqq 0, \tag{19}$$

L being a linear form of the ρ_z's whose sum of coefficients is zero. If in addition to the monotonicity the ρ_z are assumed to be convex, the following lemma, which is analogous to the lemma proved in [1], is applicable to forms of this type.

LEMMA. *Let S be the set of $\rho = \{\rho_1, \rho_2, \cdots, \rho_m\}$ for which*

$$\rho_1 \geqq \rho_2 \geqq \cdots \geqq \rho_m \geqq 0 \tag{20}$$

and

$$\Delta^2\rho_{\mu-1} = \rho_{\mu+1} - 2\rho_\mu + \rho_{\mu-1} \geqq 0 \qquad (\mu = 2, 3, \cdots, m-1). \tag{21}$$

Let $\alpha_1, \cdots, \alpha_m$ be constants such that $\sum_{\mu=1}^{m} \alpha_\mu = 0$ and put $A_i = \sum_{\mu=1}^{i} \alpha_\mu$. Then

$$L = \sum_{\mu=1}^{m} \alpha_\mu \rho_\mu \geqq 0 \qquad \text{for all } \rho \varepsilon S$$

if and only if

$$B_j = \sum_{i=1}^{j} A_i \geqq 0 \qquad \text{for } j = 1, 2, \cdots, m-1. \tag{22}$$

Moreover, if in addition to (20) and (21) strict inequality holds in (22), then $L > 0$ unless $\rho_1 = \cdots = \rho_m$.

PROOF. Writing $\alpha_\mu = A_\mu - A_{\mu-1} (\mu = 1, \cdots, m; A_0 = 0)$ and using the fact that $A_m = 0$, we find

$$L = \sum_{\mu=1}^{m} A_\mu \rho_\mu - \sum_{\mu=1}^{m} A_{\mu-1}\rho_\mu = - \sum_{\mu=1}^{m-1} A_\mu \Delta\rho_\mu.$$

Similarly,

$$\sum_{\mu=1}^{m-1} A_\mu \Delta\rho_\mu = - \sum_{\mu=1}^{m-2} B_\mu \Delta^2\rho_\mu + B_{m-1}(\rho_m - \rho_{m-1}).$$

Thus

$$L = \sum_{\mu=1}^{m-2} B_\mu \Delta^2 \rho_\mu + B_{m-1}(\rho_{m-1} - \rho_m). \tag{23}$$

Since, by hypothesis, the coefficients of all the B_μ are nonnegative, the sufficiency of (22) is clear. On the other hand, if $B_{m-1} < 0$, we could choose the ρ_μ linearly decreasing and obtain $L < 0$. If $B_j < 0$, $1 \leqq j \leqq m-2$, L could be made negative by taking, for example,

$$\rho_\mu = \begin{cases} j + 2 - \mu, & 1 \leqq \mu < j + 1, \\ 1, & j + 1 \leqq \mu \leqq m. \end{cases}$$

Thus (22) is also a necessary condition. If all the B_j are positive, then $L = 0$ implies $\Delta^2\rho_\mu = 0 (\mu = 1, \cdots, m-2)$, $\rho_{m-1} = \rho_m$. This in turn implies that $\rho_{m-2} = \rho_{m-1}, \rho_{m-3} = \rho_{m-2}, \cdots, \rho_1 = \rho_2$.

THEOREM. *For any population in which*

$$\rho_1 \geqq \rho_2 \geqq \cdots \geqq \rho_{N-1} \geqq 0,$$

$$\Delta^2\rho_{z-1} = \rho_{z+1} - 2\rho_z + \rho_{z-1} \geqq 0 \quad (z = 2, \cdots, N-2)$$

we have

$$EV_l^{(1)} \leqq EV_k^{(s)} \tag{24}$$

with equality only if $s = 1$ *or* $\rho_1 = \cdots = \rho_{N-1}$.

PROOF. There is nothing to prove if $s = 1$. If $s > 1$ we apply the above lemma (with $m = N - 1$ and L given by (17), (18) and (19)) and show that

$$B_j > 0 \qquad j = 1, 2, \cdots, N-2. \tag{25}$$

We notice that

$$\frac{N}{2} L_1 = \frac{1}{l-1}\left[\sum_{z=1}^{N-1} (N-z)\rho_z - l^2 \sum_{z=1}^{ns-1} (ns-z)\rho_{ls}\right]$$

$$\frac{N}{2} L_2 = \frac{1}{ls-1}\left[\sum_{z=1}^{N-1} (N-z)\rho_z - (ls)^2 \sum_{z=1}^{n-1} (n-z)\rho_{(ls)z}\right].$$

To prove (25) it is enough to show that the sums B_j are positive for the form $NL/2 = NL_1/2 - NL_2/2$. We compute these sums separately for $NL_1/2$, $NL_2/2$ and then take their differences. Put[4]

$$\begin{aligned} &j = \nu k + \sigma l + \lambda = (\nu s + \sigma) l + \lambda, \qquad \text{where } \nu = 0, 1, \cdots, n-1; \\ &\sigma = 0, 1, \cdots, s-1; \qquad \lambda = 0, 1, \cdots, l-1. \end{aligned} \tag{26}$$

[4] We use the Greek letters ν, σ, λ to indicate their range $n-1$, $s-1$, $l-1$, respectively; σ, λ should not be confused with the variance symbol and the parameter to be introduced in Section 5.

By elementary computations the sums $B_j^{(1)}$ for $NL_1/2$ are found to be

$$B_j^{(1)} = \frac{1}{l-1}\{\mathrm{I} - \mathrm{II}\},$$

where

$$\mathrm{I} = \sum_{i=1}^{(\nu s+\sigma)l+\lambda} \frac{i(2N-i-1)}{2}$$

$$= \tfrac{1}{6}[(\nu s+\sigma)l+\lambda][(\nu s+\sigma)l+\lambda+1][3N-(\nu s+\sigma)l-\lambda-2]$$

$$\mathrm{II} = l^2\left[l\sum_{i=1}^{\nu s+\sigma-1}\frac{i(2ns-i-1)}{2} + (\lambda+1)\frac{(\nu s+\sigma)(2ns-\nu s-\sigma-1)}{2}\right]$$

$$= \frac{l^2(\nu s+\sigma)}{6}[l(\nu s+\sigma-1)(3ns-\nu s-\sigma-1)$$

$$+ 3(\lambda+1)(2ns-\nu s-\sigma-1)].$$

Similarly the sums $B_j^{(2)}$ for $NL_2/2$ are obtained as

$$B_j^{(2)} = \frac{1}{ls-1}\{\mathrm{I} - \mathrm{III}\},$$

where

$$\mathrm{III} = (ls)^2\left[ls\sum_{i=1}^{\nu-1}\frac{i(2n-i-1)}{2} + (\sigma l+\lambda+1)\frac{\nu(2n-\nu-1)}{2}\right]$$

$$= \frac{\nu(ls)^2}{6}[ls(\nu-1)(3n-\nu-1) + 3(\sigma l+\lambda+1)(2n-\nu-1)].$$

We have to show that

$$B_j = B_j^{(1)} - B_j^{(2)} = \frac{1}{6}\frac{l}{(l-1)(ls-1)}$$

$$\cdot\left[(s-1)6\mathrm{I} - (ls-1)\frac{6\mathrm{II}}{l} + (l-1)\frac{6\mathrm{III}}{l}\right] > 0.$$

After some elementary algebra the expression in brackets is found to be a polynomial $f(\sigma)$ in σ of third degree with the following coefficients

$$\sigma^3\colon\ l^2(l-1)$$

$$\sigma^2\colon\ -3l(l-1)[(n-\nu)ls - (\lambda+1)]$$

$$\sigma^1\colon\ l\{(l-1)[3s(n-\nu)(sl-2(\lambda+1)) - sl + 3s\lambda + 2s + 1]$$

$$- 3\lambda(\lambda+1)(s-1)\} \tag{27}$$

$$\sigma^0\colon\ (s-1)\{\nu ls(l-1)(ls-1) + \lambda(\lambda+1)[3ls(n-\nu) - (\lambda+2)]\}.$$

We notice that the second derivative $f''(\sigma)$ vanishes at

$$\sigma^* = (n - \nu)s - \frac{\lambda + 1}{l}$$

which is $\geqq s - 1$ whatever be the values of ν, λ specified by (26). For any of those values $f(\sigma)$ is therefore concave between $\sigma = 0$ and $\sigma = s - 1$ so that it is enough to show $f(0) > 0$, $f(s - 1) > 0$. Now, if $\sigma = 0$ then not both ν, λ can vanish. Hence, $f(0) > 0$ follows immediately from (27). On the other hand, $f(s - 1)/(s - 1)$, after some slight rearranging, can be written as

$$\begin{aligned}\frac{f(s-1)}{s-1} &= 3(n-\nu)sl[(l-1)(l-2(\lambda+1)) + \lambda(\lambda+1)] \\ &+ l\{l(l-1)((s-1)^2 - s) + 3(s-1)(\lambda+1)(l-1-\lambda)\} \\ &+ \lambda\{3sl(l-1) - (\lambda+1)(\lambda+2)\} + l(l-1)\{2s + \nu s(ls-1) + 1\}.\end{aligned} \tag{28}$$

The expression in brackets is a polynomial of second degree in λ with a positive leading coefficient and with roots $\lambda = l - 2$, $\lambda = l - 1$. It is therefore nonnegative for $\lambda = 0, 1, \cdots, l - 1$. It is easily verified that the quantities in the three braces are nonnegative for $l > 1$, $s > 2$ and λ, ν satisfying (26). Furthermore, the last term is positive. It remains to consider (28) for the particular case $s = 2$. We have

$$\begin{aligned}f(1) \geqq\ & 6l[(l-1)(l-2(\lambda+1)) + \lambda(\lambda+1)] \\ &+ l\{3(\lambda+1)(l-1-\lambda) - l(l-1)\} \\ &+ \lambda\{6l(l-1) - (\lambda+1)(\lambda+2)\} + 5l(l-1).\end{aligned}$$

The right-hand side is a polynomial $\varphi(\lambda)$ of third degree,

$$\varphi(\lambda) = -\lambda^3 + 3(l-1)\lambda^2 - (3l^2 - 6l + 2)\lambda + l(l-1)(5l-4),$$

whose second derivative $\varphi''(\lambda)$ vanishes at $\lambda = l - 1$. It is easy to verify that $\varphi(\lambda)$ has its relative minimum at $\lambda = l - 1 - \sqrt{3}/3$. Hence $\varphi(\lambda) > 0$ for $\lambda = 0, 1, \cdots, l - 1$ follows from

$$\varphi(l-2) = \varphi(l-1) = 2l(2l-1)(l-1) > 0.$$

This completes the proof of our theorem.

For populations with serial correlation the result (24) is to be expected also on intuitive grounds; in fact, the systematic sample is spread more evenly through the population than the sample with multiple random starts which may contain elements very close together, giving about the same information. Our proof, however, does not make clear why (24) only holds for populations with a *convex* correlogram. That (24) does not generally hold for any monotone decreasing correlogram can readily be seen by trying to apply Cochran's lemma [1] to the linear form (19). It turns out that, for example, the sum of the first l coefficients of $NL/2$ is equal to

$$\frac{-l^2}{2(ls-1)}[(2n-1)s-1] < 0.$$

One might suspect that $EV_l^{(1)} \geqq EV_k^{(s)}$ for all populations with a *concave* decreasing correlogram. However, according to our theorem $EV_l^{(1)} < EV_k^{(s)}$ for the example of a linear correlogram, so that the conjecture is not generally true.

5. Asymptotic results in the case of an exponential correlogram. We assume that $\rho_z = e^{-\lambda z}(z = 1, \cdots, N-1)$ and that both l and n are large. For n, k large Cochran [1] showed that the expression in braces of (12) is approximately equal to $1 - 2/\lambda k + 2/(e^{\lambda k} - 1)$. Since the corresponding expression $1 - L_1$ in (17) is obtained by replacing k by l and n by ns, we find

$$1 - L_1 \sim 1 - \frac{2}{\lambda l} + \frac{2}{e^{\lambda l} - 1}. \tag{29}$$

On the other hand, replacing l by $k = ls$, s by 1 in the brace of (17), we obtain $1 - L_2$ of (18). Thus

$$1 - L_2 \sim 1 - \frac{2}{\lambda ls} + \frac{2}{e^{\lambda ls} - 1}.$$

Introducing $\rho = e^{-\lambda l}$, we see that the relative precision of systematic sampling over systematic sampling with multiple random starts

$$RP = \frac{EV_k^{(s)}}{EV_l^{(1)}} \sim \frac{1 + \dfrac{2}{s \log \rho} + \dfrac{2\rho^s}{1-\rho^s}}{1 + \dfrac{2}{\log \rho} + \dfrac{2\rho}{1-\rho}}$$

depends, apart from s, only on the correlation ρ of elements of a distance l apart. Clearly $\lim_{\rho \downarrow 0} RP = 1$; also, expanding numerator and denominator in power series, it is readily seen that $\lim_{\rho \uparrow 1} RP = s$. The numerical values in Table 1 show that the limit as $\rho \downarrow 0$ is approached rather slowly.

6. Concluding remark. When the statistician has a choice between systematic sampling and systematic sampling with multiple random starts, he is more

TABLE 1

Relative precision RP of systematic sampling over systematic sampling with multiple random starts for an exponential correlogram

s	ρ										
	.01	.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	1.34	1.53	1.66	1.80	1.87	1.92	1.95	1.97	1.99	2.00	2.00
5	1.56	1.98	2.34	2.92	3.43	3.88	4.25	4.55	4.76	4.92	4.99
10	1.63	2.13	2.58	3.40	4.26	5.19	6.23	7.32	8.39	9.31	9.85

likely to use the latter procedure because its variance can be estimated from the sample and the estimate is unbiased whatever be the form of the population. On the other hand, as we have seen in Section 5, systematic sampling is considerably more precise in the case of a population with an exponential correlogram. Thus, it may be worth while to try to find an estimate for the variance of systematic sampling which is at least consistent in some sense if the underlying assumption of an exponential correlogram is realized. In view of (17) or (29) this would involve estimating the correlation between the elements as well as σ^2.

REFERENCES

[1] W. G. Cochran, "Relative accuracy of systematic and stratified random samples for a certain class of populations," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 164–177.
[2] W. G. Cochran, *Sampling Techniques*, John Wiley and Sons, 1953.
[3] W. E. Deming, *Some Theory of Sampling*, John Wiley and Sons, 1950.
[4] H. L. Jones, "The application of sampling procedures to business operations," *J. Amer. Stat. Assoc.*, Vol. 50 (1955), pp. 763–774.
[5] H. L. Jones, "Investigating the properties of a sample mean by employing random subsample means," *J. Amer. Stat. Assoc.*, Vol. 51 (1956), pp. 54–83.
[6] W. G. Madow and L. H. Madow, "On the theory of systematic sampling," *Ann. Math Stat.*, Vol. 15 (1944), pp. 1–24.
[7] A. Mood, *Introduction to the Theory of Statistics*, McGraw-Hill Co., 1950.

TIGHTENED MULTI-LEVEL CONTINUOUS SAMPLING PLANS

By C. Derman,[1] S. Littauer[2, 3] and H. Solomon[2, 3]

Columbia University

1. Introduction. Industrial needs have provoked some recent studies on continuous sampling. This procedure is especially of interest when the formation of inspection lots for lot-by-lot acceptance may be impractical or artificial as in conveyor-line production, or when there is an important need for rectifying quality of product as it is manufactured.

These newer papers are best considered in the light of the earlier papers of Dodge [3] and Wald and Wolfowitz [11]. One point of departure from the Dodge type of plan has been the introduction of several levels of partial inspection with different rates of sampling in each level. Multi-level continuous sampling plans (which reduce to the Dodge plan when only one sampling level is tolerated) have been considered by Greenwood [8], Lieberman and Solomon [9], and Resnikoff [10]. A plan based on the Wald-Wolfowitz approach, a scheme essentially handled by the methodology of sequential analysis, was created and developed by Girshick about 1948 in connection with a Census Bureau problem and has only recently been reported [7]. The reader is referred to Bowker [1] for a more thorough account of continuous sampling plans.

The multi-level plan given in [9], namely MLP, allows for any number of sampling levels, subject to the provision that transitions can only occur between adjacent levels. Three generalizations of MLP, accomplished by altering the manner in which transition can occur, are analyzed in this paper. In each situation, we will make it more difficult to get to infrequent inspection than in MLP, and thus we can label these three plans as tightened plans. These three plans which will now be specifically defined obviously relate to more realistic situations for control of industrial processes. The three plans are given in language which assumes some familiarity with MLP, which is given in detail in [9].

(a) *The MLP–$r \times 1$ Plan.* We say we are in the jth sampling level if every $(1/f)^j$-th item produced is systematically sampled. If i consecutively inspected items are found clear of defects when sampling at the jth level, begin sampling at the $(j + 1)$-th level. On the other hand, if a defective item is found before this is accomplished, revert immediately to the $(j - r)$-th level, if $j > r$, or to the zero level, that is, one hundred per cent inspection if $j \leqq r$. Let inspection begin at the zero level. When $r = 1$, we have the MLP plan described in [9].

(b) *The MLP–T Plan.* This is exactly the same as the MLP–$r \times 1$ Plan, except that when a defective is encountered, we immediately revert to one hun-

Received May 24, 1956; revised September 14, 1956.

[1] Work sponsored in part by the Office of Scientific Research, U. S. Air Force.

[2] Work sponsored in part by the Office of Naval Research.

[3] Work sponsored in part by the Higgins Fund, Columbia University.

dred per cent inspection. This is obviously the tightest of the three multi-level plans considered in this paper and thus bears the label *MLP–T*.

(c) *The MLP–r* $\times$ *s Plan.* This plan follows exactly the same pattern as the *MLP–r* $\times$ 1, except that when i consecutively inspected items are found nondefective while on the jth sampling level, systematic sampling begins at level $(j + s)$. We shall consider the case $r > s$, since we are concerned only with tightened multi-level plans. If $r = s$, we are effectively using the *MLP* Plan.

2. Summary. Each of these generalizations can be appraised under the assumption of an infinite number of sampling levels or a finite number, k, of sampling levels. Under the assumption of an infinite number of allowable sampling levels, it is possible to obtain explicit relationships between the *AOQL* and the parameters of the plan for *MLP–r* $\times$ 1 and *MLP–T*. Thus it is possible to graph contours of equal *AOQL* for each of these plans under these conditions. Approximations for contours of equal *AOQL* for the *MLP–r* $\times$ *s* Plan are then easily obtained. This makes feasible the possibility of a catalogue of continuous sampling plans which contains plans having a prescribed *AOQL* and thus aids immeasurably in the choice of an appropriate plan. As is demonstrated in the next sections, the following results are obtained, assuming that the production process is in statistical control and items found defective on inspection are replaced with good items. For the *MLP–r* $\times$ 1 Plan:

$$AOQL = 1 - \left(\frac{f - f^{r+1}}{1 - f^{r+1}}\right)^{1/s}. \tag{2.1}$$

When $r = 1$, this reduces to the result previously obtained in [9]. For the *MLP–T* Plan:

$$AOQL = 1 - f^{1/s}. \tag{2.2}$$

This result can also be obtained heuristically by letting r approach infinity in *MLP–r* $\times$ 1. For the *MLP–r* $\times$ *s* Plan $(r > s)$ bounds and sometimes exact *AOQL*'s can be obtained using the previous two results. For example, if $r = 4$ and $s = 2$ and f is given, the *MLP*–2 $\times$ 1 Plan for $f' = f^2$ will be the same plan and hence have the same *AOQL*. More generally for a given f we can write

$$AOQL_{r'',s} < AOQL_{r,s} < AOQL_{r',s} \tag{2.3}$$

where r' = greatest number less than r that is a multiple of s, and r'' is the smallest number greater than r that is a multiple of s. For, if $r' < r''$, the plan associated with r'' is tighter and the added protection thus insures a better outgoing quality, i.e., a smaller *AOQL*. Under the assumption of a finite number, k, of allowable sampling levels, the *AOQ* function for *MLP–T* is obtained, and it is seen that the use of digital computers may be expedient for the computation of *AOQL* contours. This was exactly the situation, for finite levels, in [9]. The main results of the paper are obtained through the use of Markov chain techniques which are developed in Section 3. In these plans inspection, as described,

is by systematic sampling. However, the *AOQ* and *AOQL* results also hold when inspection in each level is accomplished by random sampling—i.e., in the kth level, each item in the block of f^{-k} items has probability f^k of being chosen for inspection.

3. Markov Chain Result. Let $\{X_n\}(n = 0, 1, \cdots)$ denote an irreducible recurrent positive Markov chain with states $\{E_j\}$ $(j = 0, 1, \cdots)$. Let $\{p_{ij}\}$ $(i, j = 0, 1, \cdots)$ denote the probability of transition from state E_i to E_j. It is known (see [5]), that a unique sequence $\{v_i\}$ exists such that

$$\sum_{i=0}^{\infty} v_i p_{ij} = v_j, \qquad (j = 0, 1, \cdots),$$

$$v_i > 0, \qquad (i = 0, 1, \cdots), \tag{3.1}$$

$$\sum_{i=0}^{\infty} v_i = 1.$$

The v_i's are sometimes referred to as "steady state" probabilities.

Now let $A = \{E_{j_i}\}$ be a subset of the states. Let $Y_0, Y_1, \cdots$ be successive members of $\{X_n\}$ which take on values in A. Since the chain is recurrent, infinitely many such Y's will exist with probability one. It was shown by Derman [2] that $\{Y_k\}$ $(k = 0, 1, \cdots)$ is also a Markov chain; and if $\{p'_{ij}\}$ $(E_i, E_j \varepsilon A)$ are its transition probabilities, then the solutions v'_i of

$$\sum_{E_i \varepsilon A} v'_i p'_{ij} = v'_j \qquad (E_j \varepsilon A),$$

$$v'_i > 0 \qquad (E_i \varepsilon A), \tag{3.2}$$

$$\sum_{E_i \varepsilon A} v'_i = 1$$

are given by

$$v'_i = \frac{v_i}{\sum_{E_j \varepsilon A} v_j} \qquad (E_i \varepsilon A). \tag{3.3}$$

Suppose $A_1 = \{E_j\}$ $(j = 1, 2, \cdots)$; $A_2 = \{E_j\}$ $(j = 2, 3, \cdots)$; $\cdots A_g = \{E_j\}$ $(j = g, g + 1, \cdots) \cdots$ are subsets to be considered. Let $\{Y_k(g)\}$ denote the Markov chain defined over A_g. Also let $E_j(g)$ $(j = 0, 1, \cdots)$, the states for the chain $\{Y_k(g)\}$, be a relabeling of the states $E_k(k = g, \cdots)$ by letting $j = k - g$. Finally let $p_{ij}(g)$ denote the probability of transition from state $E_i(g)$ to state $E_j(g)$ in the chain $\{Y_k(g)\}$. Our main tool is the following theorem

THEOREM. *If* $p_{ij} = p_{ij}(g)$ $(i, j = 0, \cdots; g = 1, \cdots)$, *then*

$$v_j = v_0(1 - v_0)^j \qquad (j = 1, \cdots). \tag{3.4}$$

PROOF. Let $\{v_j(g)\}$ denote the solution of (3.1) for the chain $\{Y_k(g)\}$. Since the transition probabilities, by hypothesis, are the same regardless of which

chain is under consideration, $v_i(g) = v_i$ $(i = 0, 1, \cdots)$. However, from (3.3) we have

$$v_0 = v_0(g) = \frac{v_g}{\sum_{j=g}^{\infty} v_j} = \frac{v_g}{1 - \sum_{j=0}^{g-1} v_j} \qquad (g = 1, 2, \cdots). \tag{3.5}$$

Thus by induction,

$$\begin{aligned} v_j &= v_0(1 - v_0 - \cdots - v_{j-1}) \\ &= v_0 \left[1 - v_0 - v_0 \sum_{i=1}^{j-1} (1 - v_0)^i\right] \\ &= v_0 (1 - v_0)^j \qquad (j = 1, \cdots), \end{aligned} \tag{3.6}$$

and the theorem is proved.

We shall apply the theorem in the following case. Suppose

$$\begin{aligned} p_{i,i+1} &= \alpha > 0 && (i = 0, 1, \cdots), \\ p_{i,0} &= 1 - \alpha && (i = 0, 1, \cdots, r), \\ p_{i,i-r} &= 1 - \alpha && (i > r). \end{aligned}$$

It is clear that the chain is irreducible. It also follows from a slightly modified theorem of Foster ([6], Theorem 5, p. 81) that the chain is recurrent positive if $\alpha < r/(r + 1)$. Intuitively this condition guarantees a sufficient pull to the left, thereby insuring the existence of the steady-state probabilities inherent in a recurrent positive chain. Furthermore, it is easily seen that the conditions of the theorem are satisfied so that the v_j have the form (3.4). From (3.1), $j = 0$, v_0 is determined by the following equation

$$(1 - \alpha)\left\{\frac{1 - (1 - v_0)^{r+1}}{v_0}\right\} = 1, \tag{3.7}$$

and thus any v_j can be obtained.

4. Application to MLP-r $\times$ 1 infinite-level plan. The multilevel plans can now be studied from the point of view of a Markov chain $\{X_n\}$ and the results in Section 3 employed. We let $E_{jm}(j = 0, 1, \cdots; m = 0, \cdots, i - 1)$ denote the state of such a chain where we say that X_n is in state E_{jm} if just after the nth item has been inspected, the process is in the jth sampling level (i.e., every (f^{-j})th item inspected) and m nondefectives have been observed successively while in the jth level. Suppose the process is in a state of control such that p is the probability of a defective being produced. The transition probabilities are then given by

$$P(E_{jm} \to E_{j,m+1}) = 1 - p = q$$
$$(j = 0, 1, \cdots; m = 0, 1, \cdots, i - 2)$$

$$P(E_{j,i-1} \to E_{j+1,0}) = q \qquad (j = 0, 1, \cdots),$$
$$P(E_{jm} \to E_{j-r,0}) = p \qquad (j = r, \cdots),$$
$$P(E_{jm} \to E_{00}) = p \qquad (j = 1, \cdots, r-1). \tag{4.1}$$

The chain is easily seen to be irreducible. From Foster's theorem it is seen to be recurrent positive if $q^i < r/(r+1)$. We shall assume $q < r/(r+1)$ for the present. Now let $A = \{E_{j0}\}$ be a subset of the states and let $\{Y_k\}$ denote the chain defined over it. The chain is of the form of the special case considered in section 3 with $\alpha = q^i$. Let $\{v'_j\}$ and $\{v_{jm}\}$ denote the steady-state probabilities of the chains $\{Y_k\}$ and $\{X_n\}$, respectively. Using (3.1), (3.5) and (4.1) it follows that

$$v_{jm} = \frac{1-q}{1-q^i} v'_j q^m \qquad (m = 0, 1, \cdots, i-1; j = 0, 1, \cdots). \tag{4.2}$$

For from (3.1)

$$v_{jm} = v_{j0} q^m \qquad (m = 0, \cdots, i-1; j = 0, 1, \cdots),$$

and from (3.5)

$$v'_j = \frac{v_{j0}}{\sum_{k=0}^{\infty} v_{k0}} \qquad (j = 0, 1, \cdots).$$

Hence,

$$v_{jm} = \sum_{k=0}^{\infty} v_{k0} v'_j q^m \qquad (j = 0, 1, \cdots);$$

but summing over j and m we get, since $\sum_{j,m} v_{jm} = 1$,

$$\sum_{k=0}^{\infty} v_{k0} = \frac{1-q}{1-q^i}.$$

From (4.2) it is clear that v'_j is the sum of the steady-state probabilities of being in the jth level of sampling. Also from (3.4)

$$v'_j = v'_0 (1 - v'_0)^j \qquad (j = 1, 2, \cdots), \tag{4.3}$$

where v'_0 is given by (3.7) with $\alpha = q^i$; namely,

$$(1 - q^i)\left[\frac{1 - (1 - v'_0)^{r+1}}{v'_0}\right] = 1,$$

where as previously remarked, v'_0 is the probability of being in one hundred per cent inspection.

Now that we have expressions for the steady-state probabilities, we proceed with the derivation of the AOQ functions and the $AOQL$. Let $h(X_n) = f^{-j}$ for

$X_n = E_{jm}$. It is easily verified that the reciprocal of the average fraction inspected after n inspections is

$$F_n^{-1} = \frac{1}{n} \sum_{\nu=1}^{n} h(X_\nu). \tag{4.4}$$

It follows from the Birkhoff ergodic theorem, applicable for stationary Markov chains of the type considered here (see Doob [1], p. 460), that

$$F^{-1} = \lim_{n \to \infty} F_n^{-1} = \sum_{j=0}^{\infty} f^{-j} \sum_{m=0}^{j-1} v_{jm}. \tag{4.5}$$

Now F^{-1} denotes the reciprocal of the average fraction inspected for all sequences (except for a set having probability 0); for let $t_k = \sum_{m=1}^{k} h(X_m)$ = number of items produced during the first k inspections. Formula (4.5) says that $k/t_k \to F$ as $k \to \infty$. Let $t_k < t < t_{k+1}$. Then since k = number of items inspected in the first t items produced, the inequalities

$$\frac{k}{t_{k+1}} < \frac{k}{t} \leqq \frac{k}{t_k}$$

imply that $\lim_{t \to \infty} k/t \to F$ with probability 1.

If $q^i \geqq r/(r+1)$, it can be shown more directly that $F^{-1} = \infty$ with probability 1. If v_0' exists and is positive, it follows from the theory of recurrent Markov chains that $q^i < r/(r+1)$. Thus since $0 < f < 1$, we have from (4.2), (4.3), (4.5) and the last remark that

$$\begin{aligned} F^{-1} &= v_0' \left(\frac{1}{1 - \dfrac{1 - v_0'}{f}} \right), && \text{when } (f > 1 - v_0'), \\ &= \infty, && \text{otherwise.} \end{aligned} \tag{4.6}$$

Hence since it can easily be shown that $AOQ = p(1 - F)$, we have

$$\begin{aligned} AOQ &= (1 - q) \left(\frac{1 - f}{f} \right) \frac{1 - v_0'}{v_0'}, && \text{when } (f > 1 - v_0'), \\ &= 1 - q, && \text{otherwise.} \end{aligned} \tag{4.7}$$

Now suppose it is true that the AOQ is an increasing function of q as long as $f > 1 - v_0'$. Then from (4.7) it would follow that

$$AOQL = 1 - q_0, \tag{4.8}$$

where q_0 is the value of q such that $f = 1 - v_0'$. From (3.7) with $\alpha = q^i$, it is easily established that

$$q_0 = \left(\frac{f - f^{r+1}}{1 - f^{r+1}} \right)^{1/i},$$

so that

$$AOQL = 1 - \left(\frac{f - f^{r+1}}{1 - f^{r+1}}\right)^{1/i}. \tag{4.9}$$

We now show that the AOQ is an increasing function of q as long as

$$q < \left(\frac{f - f^{r+1}}{1 - f^{r+1}}\right)^{1/i} \qquad \text{(i.e., } f > 1 - v_0'\text{)}.$$

Let

$$\varphi(q) = \left(\frac{f}{1 - f}\right) AOQ = (1 - q)\frac{1 - v_0'}{v_0'}$$

and

$$V(q) = \frac{1 - v_0'}{v_0'}.$$

Then

$$\frac{d\varphi(q)}{dq} = -V(q) + (1 - q)\frac{dV(q)}{dq}. \tag{4.10}$$

It is necessary to show that the right-hand side of (4.10) is positive or

$$\frac{V(q)}{(1 - q)\dfrac{dV(q)}{dq}} \leqq 1. \tag{4.11}$$

But, using (3.7) with $\alpha = q^i$,

$$\frac{dV(q)}{dq} = \left(-\frac{1}{v_0'^2}\right)\left\{\frac{iq^{i-1}v_0'}{(1 - q^i)^2\left[(r + 1)(1 - v_0')^r - \dfrac{1}{(1 - q^i)}\right]}\right\}. \tag{4.12}$$

Thus the left side of (4.11) becomes

$$\frac{-(1 - q^i)\left[(r + 1)(1 - v_0')^{r+1}(1 - q^i) - (1 - v_0')\right]}{iq^{i-1}(1 - q)}. \tag{4.13}$$

From (3.7) it follows that $(1 - v_0')^{r+1} = [(1 - v_0') - q^i]/(1 - q^i)$. Hence (4.13) becomes

$$-\frac{q}{i}\left(\frac{1 - q^i}{1 - q}\right)\left[\frac{(1 - v_0')r}{q^i} - (r + 1)\right]. \tag{4.14}$$

But from (3.7),

$$q^i = (1 - v_0')\frac{1 - (1 - v_0')r}{1 - (1 - v_0')^{r+1}} \leqq 1 - v_0'.$$

Hence

$$\frac{(1 - v_0')r}{q^i} \geqq r,$$

and the smallest value over the range $f > 1 - v_0'$ which the bracket factor in (4.14) can take is minus one. Thus the largest value that (4.14) can reach is

$$\frac{(1 - q^i)}{1 - q}\left(\frac{q}{i}\right). \tag{4.15}$$

But

$$\frac{1 - q^i}{1 - q}\left(\frac{q}{i}\right) = \frac{q + q^2 + \cdots + q^i}{i} < 1.$$

This proves (4.11).

5. The MLP–T Plan. We consider first an infinite number of sampling levels. Let E_{jm} be as in the previous section. The transition probabilities are now

$$P(E_{jm} \to E_{j,m+1}) = q \qquad (j = 0, 1, \cdots ; 0 < m \leqq i - 2),$$

$$P(E_{j,i-1} \to E_{j+1,0}) = q \qquad (j = 0, 1, \cdots),$$

$$P(E_{jm} \to E_{\infty}) = 1 - q \qquad (\text{for all } j, m).$$

Of course, $0 < q < 1$.

It can be shown in this case that

$$v_{jm} = pq^{ji+m} \qquad (j = 0, 1, \cdots ; m = 0, \cdots , i - 1),$$

and as before that

$$F^{-1} = \sum_{jm} f^{-i} v_{jm} = \frac{1 - q^i}{1 - \frac{q^i}{f}} \qquad (f > q^i),$$

$$= \infty \qquad (f \leqq q^i);$$

and

$$AOQ = \frac{(1 - q)\, q^i}{1 - q^i}\left(\frac{1 - f}{f}\right) \qquad (f > q^i),$$

$$= 1 - q \qquad (f \leqq q^i).$$

It can easily be shown that AOQ is an increasing function of q for $0 < q^i < f$.

Hence,

$$AOQL = 1 - f^{1/i}.$$

Now let the number of sampling levels, k, be finite. For this case we need only modify the function $h(X_n)$ such that

$$h(X_n) = f^{-j} \quad \text{when} \quad X_n = E_{jm} \qquad (j \leqq k),$$
$$= f^{-k} \quad \text{when} \quad X_n = E_{jm} \qquad (j > k),$$

where here we persist with the notation E_{jm} as if the $k = \infty$ plans are in effect. In similar fashion we have

$$F^{-1} = p \sum_{j=0}^{k-1} \sum_{m=0}^{i-1} f^{-j} q^{ji+m} + p \sum_{j=k}^{\infty} \sum_{m=0}^{i-1} f^{-k} q^{ji+m}$$

$$= (1 - q^i) \frac{1 - (q^i/f)^k}{1 - q^i/f} + (q^i/f)^k.$$

For $k = 1$, we have the Dodge Plan, and get the following result as in [3]:

$$F^{-1} = \frac{f}{f + q^i(1 - f)}.$$

For $k = 2$,

$$F^{-1} = 1 + q^i \left(\frac{1 - f}{f}\right) + q^{2i} \left(\frac{1 - f}{f^2}\right).$$

In order to obtain $AOQL$ contours for this situation, as for higher values of k, the use of digital computers would be expedient.

REFERENCES

[1] A. H. Bowker, "A survey of continuous sampling plans," *Proceedings of the Third-Berkeley Symposium on Mathematical Statistics and Probability*, Vol. V, University of California Press, 1956, pp. 75–86.

[2] C. Derman, "Some contributions to the theory of denumerable Markov chains," *Trans. Amer. Math. Soc*, Vol. 79, No. 2 (1955), pp. 541–555.

[3] H. F. Dodge, "A sampling inspection plan for continuous production," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 264–279.

[4] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, 1953.

[5] W. Feller, *Probability Theory and Its Applications*, John Wiley and Sons, 1950.

[6] F. G. Foster, "Markoff chains with an enumerable number of states and a class of cascade processes," *Proc. Cambridge Phil. Soc.*, Vol. 47 (1951), pp. 77–85.

[7] M. A. Girshick, "A sequential inspection plan for quality control," Technical Report No. 16, Applied Mathematics and Statistics Laboratory, Stanford University, 1954.

[8] J. A. Greenwood, "A continuous sampling plan and its operating characteristics," Bureau of Ordnance, Navy Dept., Washington, D. C., unpublished memorandum.

[9] G. LIEBERMAN AND H. SOLOMON, "Multi-level continuous sampling plans," *Ann. Math. Stat.*, Vol. (26), No. 4 (1955), pp. 686–704.

[10] G. RESNIKOFF, "Some modifications of the Lieberman-Solomon multi-level continuous sampling plan, MLP," Technical Report No. 26, Applied Mathematics and Statistics Laboratory, Stanford University, 1956.

[11] A. WALD AND J. WOLFOWITZ, "Sampling inspection plans for continuous production which insure a prescribed limit on the outgoing quality," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 30–49.

ON SOME CHARACTERIZATION PROBLEMS CONNECTED WITH LINEAR STRUCTURAL RELATIONS

By R. G. Laha

Indian Statistical Institute, Calcutta

1. Introduction. The problems concerned with the characterizations of the distribution laws of random variables when they are connected by a linear structural relation seem to originate from the stimulating problem first proposed by Ragnar Frisch before the Oxford Conference of the Econometric Society in 1936. His problem may be stated as follows. Let x_0 and x_1 be two observable random variables connected by a linear structural set up,

$$x_0 = a_0\xi + \eta_0 ,$$

$$x_1 = a_1\xi + \eta_1 ,$$

where ξ, η_0 and η_1 are mutually independent random variables, and a_0 and a_1 are some unknown constants. What are the conditions on the distribution laws of the random variables ξ, η_0 and η_1 under which the regression of x_0 on x_1 and also that of x_1 on x_0 is linear, irrespective of the values of the constants a_0 and a_1?

A partial solution to the problem of Ragnar Frisch was given by Allen [1] by proving that if the first two moments of η_0 and all the moments of ξ and η_1 exist, then a necessary and sufficient condition for the regression of x_0 on x_1 to be linear irrespective of the values of the constants a_0 and a_1 is that both ξ and η_1 are normally distributed. A more general theorem was proved later independently by Rao [10], [11] and Fix [4] as a complete solution to the problem of Ragnar Frisch. Rao-Fix's theorem may be stated as follows: Let ξ, η_0 and η_1 be three mutually independent proper random variables each having a finite expectation. Then a necessary and sufficient condition for the regression of $x_0 = a_0\xi + \eta_0$ on $x_1 = a_1\xi + \eta_1$ to be linear for some $a_0 \neq 0$ and for all a_1 contained in a closed interval is that both ξ and η_1 should belong to a class of stable laws with finite expectation.

Recently the author [6] has obtained a generalization of Rao-Fix's theorem in a new direction, replacing the condition of stochastic independence of η_0 and η_1 by the weaker assumption that the regression of η_0 on η_1 is linear. The author [5] has also obtained a characterization of the normal law from the consequence of the linearity of multiple regression of one random variable on several others, when the variables are connected by a linear structural relation as in the case of the bifactor theory of Spearman. Several analogous characterization problems connected with linear structural relations have also been solved recently by Ferguson [3]. In the present paper we shall consider some generalizations of these problems in various directions. In Section 4, some theorems on

Received June 14, 1956.

the dependent error variables (Theorems 4.1 and 4.2) are deduced and a general theorem on a higher dimensional linear structure is proved in the subsequent section (Theorem 5.1).

2. Definitions and assumptions. Throughout the present investigation we shall confine our attention to a set of observable random variables $x_0, x_1, \cdots, x_n$ having the following linear structural set up:

$$(2.1)\qquad \begin{cases} x_0 = a_{01}\xi_1 + a_{02}\xi_2 + \cdots + a_{0p}\xi_p + \eta_0, \\ x_1 = a_{11}\xi_1 + a_{12}\xi_2 + \cdots + a_{1p}\xi_p + \eta_1, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n = a_{n1}\xi_1 + a_{n2}\xi_2 + \cdots + a_{np}\xi_p + \eta_n, \end{cases}$$

where $\xi_1, \xi_2, \cdots, \xi_p$ are usually called the latent or hypothetical variables and $\eta_0, \eta_1, \cdots, \eta_n$ the error variables of the linear structural relation, and a_{ij}'s are a certain fixed set of constants.

Now using the notations of vectors and matrices, the equation (2.1) may be rewritten as

$$(2.2)\qquad (x_0 : x) = \xi(\alpha_0' : A') + (\eta_0 : \eta)$$

where

$$x = (x_1, x_2, \cdots, x_n), \qquad \xi = (\xi_1, \xi_2, \cdots, \xi_p),$$

$$\alpha_0 = (a_{01}, a_{02}, \cdots, a_{0p}), \qquad \eta = (\eta_1, \eta_2, \cdots, \eta_n),$$

and $A = (a_{ij})_{i=1,2,\cdots n;\, j=1,2,\cdots p}$ and α_0' and A' denote respectively the transposes of α_0 and A.

We shall now make the following assumptions on the distributions of the random variables concerned.

Assumption 1. The conditional distribution of x_0 for fixed $x_1, x_2, \cdots, x_n$ is assumed to exist, wherever needed.

Assumption 2. The set of random variables $\xi_1, \xi_2, \cdots, \xi_p$ is distributed independently of the set of error variables $\eta_0, \eta_1, \cdots, \eta_n$.

Assumption 3. The latent variables $\xi_1, \xi_2, \cdots, \xi_p$ are mutually independent proper random variables each having a finite expectation (which is assumed to be zero without any loss of generality in the proof) as well as a finite variance $\sigma_j^2, j = 1, 2, \cdots, p$. Similarly the error variables $\eta_0, \eta_1, \cdots, \eta_n$ are mutually independent (not necessarily proper) random variables each having a finite expectation (which is also assumed to be zero) as well as a finite variance δ_k^2, $k = 0, 1, \cdots, n$. Let the dispersion matrices of the random vectors ξ and η be denoted by Σ and Δ respectively such that

$$(2.3)\qquad \Sigma = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_p^2 \end{pmatrix} \quad \text{and} \quad \Delta = \begin{pmatrix} \delta_1^2 & & & \\ & \delta_2^2 & & \\ & & \ddots & \\ & & & \delta_n^2 \end{pmatrix}.$$

Now it should be noted that some or all of the elements of the matrix Δ may be zero.

But in Section 4, where some results concerned with dependent error variables are obtained for the special case of the above structure with $p = 1$ and $n \geqq 2$, it is assumed that all the random variables are proper and have only finite expectations and further the multiple regression of η_0 on $\eta_1, \eta_2, \cdots, \eta_n$ is linear.

The role of these assumptions is to ensure the existence of the expectation and the variance of the conditional distribution of x_0 for fixed $x_1, x_2, \cdots, x_n$ which we denote by $E(x_0 \mid x_1, x_2, \cdots, x_n)$ and $V(x_0 \mid x_1, x_2, \cdots, x_n)$ respectively.

3. Some lemmas. We give below some lemmas which are useful in proving the theorems in the subsequent sections.

LEMMA 3.1. *Let $x_0, x_1, \cdots, x_n$ be a set of $n + 1$ proper random variables each having a finite expectation (which is assumed to be zero without any loss of generality) as well as a finite variance. Then the necessary and sufficient conditions for*

$$(3.1) \qquad \begin{cases} E(x_0 \mid x_1, x_2, \cdots, x_n) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n, \\ V(x_0 \mid x_1, x_2, \cdots, x_n) = \sigma_0^2 \quad \text{a.e.}, \end{cases}$$

are that the equations

$$(3.2) \qquad \begin{cases} \left[\dfrac{\partial \varphi(t_0, t_1, \cdots, t_n)}{\partial t_0}\right]_{t_0=0} = \displaystyle\sum_{j=1}^{n} \beta_j \frac{\partial \varphi(0, t_1, \cdots, t_n)}{\partial t_j} \\ \left[\dfrac{\partial^2 \varphi(t_0, t_1, \cdots, t_n)}{\partial t_0^2}\right]_{t_0=0} = -\sigma_0^2 \varphi(0, t_1, \cdots, t_n) + \displaystyle\sum_{j,k=1}^{n} \beta_j \beta_k \frac{\partial^2 \varphi(0, t_1, \cdots, t_n)}{\partial t_j \partial t_k} \end{cases}$$

are to be satisfied for all real $t_1, t_2, \cdots, t_n$ where $\varphi(t_0, t_1, \cdots, t_n)$ and $\varphi(0, t_1, \cdots, t_n)$ represent respectively the characteristic functions of the distributions of $(x_0, x_1, \cdots, x_n)$ and $(x_1, x_2, \cdots, x_n)$ and further $\beta_1, \beta_2, \cdots, \beta_n$ and $\sigma_0^2 > 0$ are arbitrary constants.

When the random variables $x_0, x_1, \cdots, x_n$ satisfy the relations in (3.1), we say that the multiple regression of x_0 on $x_1, x_2, \cdots, x_n$ is linear and that the conditional distribution of x_0 for fixed $x_1, x_2, \cdots, x_n$ is homoscedastic.

PROOF: To prove that the conditions are necessary, we can easily verify that

$$\left[\frac{\partial \varphi(t_0, t_1, \cdots, t_n)}{\partial t_0}\right]_{t_0=0} = E\left\{iE(x_0 \mid x_1, x_2, \cdots, x_n) \exp\left(i \sum_{j=1}^{n} t_j x_j\right)\right\}$$

$$= \sum_{j=1}^{n} \beta_j E\left\{ix_j \exp\left(i \sum_{j=1}^{n} t_j x_j\right)\right\}$$

$$= \sum_{j=1}^{n} \beta_j \frac{\partial \varphi(0, t_1, \cdots, t_n)}{\partial t_j}.$$

Similarly

$$\left.\frac{\partial^2\varphi(t_0, t_1, \cdots, t_n)}{\partial t_0^2}\right]_{t_0=0} = -E\left\{E(x_0^2 \mid x_1, x_2, \cdots, x_n) \exp\left(i\sum_{j=1}^{n} t_j x_j\right)\right\}$$

$$= -E\left\{\left(\sigma_0^2 + \sum_{j,k=1}^{n} \beta_j \beta_k x_j x_k\right) \exp\left(i\sum_{j=1}^{n} t_j x_j\right)\right\}$$

$$= -\sigma_0^2 \varphi(0, t_1, \cdots, t_n) + \sum_{j,k=1}^{n} \beta_j \beta_k \frac{\partial^2\varphi(0, t_1, \cdots, t_n)}{\partial t_j \partial t_k}.$$

To prove the sufficiency of the conditions, we note simply that (3.2) may be rewritten as

$$E\left[\left\{E(x_0 \mid x_1, x_2, \cdots, x_n) - \sum_{j=1}^{n} \beta_j x_j\right\} \exp\left(i\sum_{j=1}^{n} t_j x_j\right)\right] = 0$$

and

$$E\left[\left\{E(x_0^2 \mid x_1, x_2, \cdots, x_n) - \sigma_0^2 - \sum_{j,k=1}^{n} \beta_j \beta_k x_j x_k\right\} \exp\left(i\sum_{j=1}^{n} t_j x_j\right)\right] = 0.$$

Then from the uniqueness theorem of Fourier transforms of functions of bounded variation, (3.1) follows immediately.

For the special case of $n = 1$, this reduces to the lemma proved independently by Rao [9] and Rothschild and Mourier [12].

LEMMA 3.2. *Let $x_0, x_1, \cdots, x_n$ be a set of $n + 1$ proper random variables each having a finite expectation (which is assumed to be zero). Then the necessary and sufficient condition for*

$$E(x_0 \mid x_1, x_2, \cdots, x_n) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad \text{a.e.}$$

is that the equation

$$\left.\frac{\partial\varphi(t_0, t_1, \cdots, t_n)}{\partial t_0}\right]_{t_0=0} = \sum_{j=1}^{n} \beta_j \frac{\partial\varphi(0, t_1, \cdots, t_n)}{\partial t_j}$$

is to be satisfied for all real values of $t_1, t_2, \cdots, t_n$.

This lemma has been already proved independently by the author [5] and Ferguson [3].

LEMMA 3.3. *Let $x_1, x_2, \cdots, x_n$ be n independent proper random variables and let further $\varphi_j(t)$ denote the characteristic function of the distribution of x_j, $j = 1, 2, \cdots, n$. If now the functions $\varphi_j(t)$ satisfy the equation*

$$\prod_{j=1}^{n} \{\varphi_j(t)\}^{\alpha_j} = e^{Q(t)},$$

for all real t in a certain neighbourhood of the origin $|t| < \delta (\delta > 0)$, where α_j's are some positive numbers and $Q(t)$ a quadratic polynomial in t, then each x_j follows normal distribution.[1]

[1] The proof of this lemma is given in A. A. Zinger and Yu. V. Linnik [13].

This lemma may be regarded as an analytical extension of Cramér's theorem on the normal law and has been proved by Linnik [8]. The proof of this lemma has been given by the author in [7].

4. Some results for the case of dependent error variables. We shall now obtain some results connected with dependent error variables for the special case of the above linear structure when $p = 1$ and $n \geqq 2$.

THEOREM 4.1. *Let the observable random variables $x_j (j = 0, 1, \cdots, n)$ have the linear structural set-up $x_j = a_j\xi + \eta_j$ where the a_j's are fixed nonzero constants and further ξ and $\eta_0, \eta_1, \cdots, \eta_n$ are proper random variables each having a finite expectation (which is assumed to be zero without any loss of generality) such that*

(i) *ξ is distributed independently of $(\eta_0, \eta_1, \cdots, \eta_n)$*

(ii) *$E(\eta_0 \mid \eta_1, \eta_2, \cdots, \eta_n) = \sum_{j=1}^{n} \beta_j'\eta_j$, the β_j''s being a set of constants, then the multiple regression of x_0 on $x_1, x_2, \cdots, x_n$ is always linear, whenever the relation $a_0 = \sum_{j=1}^{n} a_j\beta_j'$ is satisfied.*

PROOF. Let $\varphi(t_0, t_1, \cdots, t_n)$; $\varphi_0(t_0, t_1, \cdots, t_n)$ and $\Phi(t)$ represent the characteristic functions of the distributions of $(x_0, x_1, \cdots, x_n)$; $(\eta_0, \eta_1, \cdots, \eta_n)$ and ξ respectively.

Then we can write

$$\begin{aligned} \varphi(t_0, t_1, \cdots t_n) &= E[\exp\,(i\textstyle\sum_{j=0}^{n} t_jx_j)] \\ &= \Phi\left(\textstyle\sum_{j=0}^{n} a_jt_j\right)\varphi_0(t_0, t_1, \cdots, t_n). \end{aligned} \tag{4.1}$$

Again since it is given that $E(\eta_0 \mid \eta_1, \eta_2, \cdots, \eta_n) = \sum_{j=1}^{n} \beta_j'\eta_j$, by applying Lemma 3.2, we get easily

$$\frac{\partial\varphi_0(t_0, t_1, \cdots, t_n)}{\partial t_0}\Bigg]_{t_0=0} = \sum_{j=1}^{n} \beta_j' \frac{\partial\varphi_0(0, t_1, \cdots, t_n)}{\partial t_j}. \tag{4.2}$$

Now differentiating both sides of the equation (4.1) with respect to t_0 and then putting $t_0 = 0$ and finally using the equation (4.2), we have

$$\begin{aligned} \frac{\partial\varphi(t_0, t_1, \cdots, t_n)}{\partial t_0}\Bigg]_{t_0=0} &= a_0\Phi'\left(\sum_{j=1}^{n} a_jt_j\right)\varphi_0(0, t_1, \cdots, t_n) \\ &\quad + \sum_{j=1}^{n} \beta_j'\Phi\left(\sum_{j=1}^{n} a_jt_j\right)\frac{\partial\varphi_0(0, t_1, \cdots, t_n)}{\partial t_j}. \end{aligned} \tag{4.3}$$

Again putting $t_0 = 0$ on both sides of the equation (4.1) and then differentiating both sides with respect to $t_j, j = 1, 2, \cdots n$ we get

$$\begin{aligned} \frac{\partial\varphi(0, t_1, \cdots, t_n)}{\partial t_j} &= a_j\Phi'\left(\sum_{j=1}^{n} a_jt_j\right)\varphi_0(0, t_1, \cdots, t_n) \\ &\quad + \Phi\left(\sum_{j=1}^{n} a_jt_j\right)\frac{\partial\varphi_0(0, t_1, \cdots, t_n)}{\partial t_j} \qquad j = 1, 2, \cdots n. \end{aligned} \tag{4.4}$$

Now it is given that $a_0 = \sum_{j=1}^{n} a_j\beta_j'$; hence substituting this value of a_0 in

(4.3) and finally comparing with (4.4), it is easy to obtain

$$\frac{\partial \varphi(t_0, t_1, \cdots, t_n)}{\partial t_0}\Big]_{t_0=0} = \sum_{j=1}^{n} \beta_j' \frac{\partial \varphi(0, t_1, \cdots, t_n)}{\partial t_j}. \tag{4.5}$$

Then the proof follows at once using Lemma 3.2 to the equation (4.5).

THEOREM 4.2. *With the same notations and assumptions as used in Theorem* 4.1 *together with the additional assumptions*

(iii) *the variables* $\eta_1, \eta_2, \cdots, \eta_n$ *are mutually independent.*

(iv) *the constants* a_j *satisfy the relation* $a_0 \neq \sum_{j=1}^{n} a_j\beta_j'$,

the necessary and sufficient condition for the multiple regression of x_0 *on* $x_1, x_2, \cdots, x_n$ *to be linear* ($n \geqq 2$) *is that* ξ *and each of* $\eta_1, \eta_2, \cdots, \eta_n$ *is normally distributed.*

PROOF.

Necessity. Let us suppose that $E(x_0 \mid x_1, x_2, \cdots, x_n) = \sum_{j=1}^{n} \beta_j x_j$. Then using Lemma 3.2, we have

$$\frac{\partial \varphi(t_0, t_1, \cdots, t_n)}{\partial t_0}\Big]_{t_0=0} = \sum_{j=1}^{n} \beta_j \frac{\partial \varphi(0, t_1, \cdots, t_n)}{\partial t_j}. \tag{4.6}$$

Next using the equations (4.3), (4.4) and (4.6) together and noting that $\eta_1, \eta_2, \cdots, \eta_n$ are mutually independent random variables, we get after a little algebraic simplification,

$$\begin{aligned}(a_0 - \textstyle\sum_{j=1}^{n} a_j\beta_j)\Phi'(\textstyle\sum_{j=1}^{n} a_jt_j)\textstyle\prod_{j=1}^{n} \varphi_j(t_j) \\ = \textstyle\sum_{j=1}^{n} (\beta_j - \beta_j')\cdot\Phi(\textstyle\sum_{j=1}^{n} a_jt_j)\varphi_j'(t_j)\cdot\textstyle\prod_{k\neq j} \varphi_k(t_k),\end{aligned} \tag{4.7}$$

where $\varphi_j(t_j)$ represents the characteristic function of the distribution of η_j; $j = 1, 2, \cdots, n$.

It can be easily shown that under the conditions of the theorem, neither $a_0 - \sum_{j=1}^{n} a_j\beta_j$ nor any of $\beta_j - \beta_j'$ $j = 1, 2, \cdots n$ in the equation (4.7) can be equal to zero. Putting $t_k = 0$ for all $k \neq j$ in (4.7) and noting that $\varphi_j'(t_j)]_{t_j=0} = 0$ for $j = 1, 2, \cdots, n$ we get

$$(a_0 - \textstyle\sum_{j=1}^{n} a_j\beta_j)\Phi'(a_jt_j)\varphi_j(t_j) = (\beta_j - \beta_j')\Phi(a_jt_j)\varphi_j'(t_j), \quad j = 1, 2, \cdots, n. \tag{4.8}$$

Let us now suppose that $\beta_j - \beta_j' = 0$ for some j, but $a_0 - \sum_{j=1}^{n} a_j\beta_j \neq 0$. In this case the equation (4.8) gives

$$\Phi'(a_jt_j)\varphi_j(t_j) = 0. \tag{4.9}$$

But since the characteristic function $\varphi(t)$ is continuous for all real t and equal to unity at the origin, in a suitably chosen neighbourhood of the origin, we have always $\varphi_j(t_j) \neq 0$. Thus it follows that for all t_j in that neighbourhood of the origin $\Phi'(a_jt_j) = 0$, leading to the conclusion that the distribution of ξ is improper.

Proceeding in the same way it can be shown that if $\beta_j - \beta_j' \neq 0$ for any j, whereas $a_0 - \sum_{j=1}^{n} a_j\beta_j = 0$, the distribution of the corresponding η_j is im-

proper. Thus both these cases contradict the conditions of the theorem. Now the only alternative left is when $a_0 - \sum_{j=1}^{n} a_j\beta_j$ and each of $\beta_j - \beta_j' j = 1, 2, \cdots, n$ vanish simultaneously. But in this case we have $a_0 = \sum_{j=1}^{n} a_j\beta_j'$, which is also contrary to the conditions of the theorem.

Now restricting the values of $t_1, t_2, \cdots, t_n$ to a suitably chosen neighbourhood of the origin such that each of the factors occurring in the product

$$\Phi(\textstyle\sum_{j=1}^{n} a_j t_j)\prod_{j=1}^{n} \varphi_j(t_j)$$

is different from zero, we divide both sides of the equation (4.7) by this expression and thus obtain,

$$(a_0 - \textstyle\sum_{j=1}^{n} a_j\beta_j)\theta'(\sum_{j=1}^{n} a_j t_j) = \sum_{j=1}^{n} (\beta_j - \beta_j')\theta_j'(t_j) \tag{4.10}$$

where

$$\theta(t) = \ln\Phi(t) \quad \text{and} \quad \theta_j(t) = \ln\varphi_j(t), \quad j = 1, 2, \cdots, n.$$

Next putting $t_3 = t_4 = \cdots = t_n = 0$ in (4.10), we get

$$(a_0 - \textstyle\sum_{j=1}^{n} a_j\beta_j)\theta'(a_1t_1 + a_2t_2) = (\beta_1 - \beta_1')\theta_1'(t_1) + (\beta_2 - \beta_2')\theta_2'(t_2). \tag{4.11}$$

Then putting successively $t_1 = 0$ and $t_2 = 0$ in (4.11) and noting that

$$a_0 - \textstyle\sum_{j=1}^{n} a_j\beta_j \neq 0,$$

we get easily

$$\theta'(a_1t_1 + a_2t_2) = \theta'(a_1t_1) + \theta'(a_2t_2). \tag{4.12}$$

But $\theta'(t)$ being continuous in t, it at once follows from the equation (4.12) that $\theta'(t)$ is a linear function of t and hence $\theta(t)$ is a quadratic polynomial in t, thus establishing the normality of the variable ξ. Then the normality of the remaining variables η_j ; $j = 1, 2, \cdots, n$ follows simply from the equation (4.8).

Sufficiency. Let σ^2 denote the variance of the random variable ξ and δ_j^2 that for η_j $j = 0, 1, 2, \cdots, n$.

Under the conditions of the theorem, we get on using the equation (4.3)

$$\left.\frac{\partial\varphi(t_0, t_1, \cdots, t_n)}{\partial t_0}\right]_{t_0=0} = -\left[\sum_{j=1}^{n} (a_0 a_j\sigma^2 + \beta_j'\delta_j^2)t_j\right]\Phi\left(\sum_{j=1}^{n} a_j t_j\right)\prod_{j=1}^{n}\varphi_j(t_j) \tag{4.13}$$

where

$$\Phi(t) = e^{-\sigma^2t^2/2} \quad \text{and} \quad \varphi_j(t_j) = e^{-\delta_j^2t_j^2/2}; \quad j = 1, 2, \cdots, n.$$

Similarly we get, on using the equation (4.4)

$$(4.14)\qquad \frac{\partial\varphi(0, t_1, \cdots, t_n)}{\partial t_j} = -\left[a_j\left(\sum_{k=1}^{n} a_k t_k\right)\sigma^2 + \delta_j^2 t_j\right]\Phi\left(\sum_{j=1}^{n} a_j t_j\right)\prod_{j=1}^{n}\varphi_j(t_j),$$
$$j = 1, 2, \cdots n.$$

Thus using (4.13) and (4.14) together, we may write

$$(4.15)\qquad \frac{\partial\varphi(t_0, t_1, \cdots, t_n)}{\partial t_0}\Big]_{t_0=0} = \sum_{j=1}^{n}\beta_j\frac{\partial\varphi(0, t_1, \cdots, t_n)}{\partial t_j},$$

where the constants β_j are to be determined from the system of equations

$$\beta_1(a_1a_j\sigma^2) + \cdots + \beta_j(a_j^2\sigma^2 + \delta_j^2) + \cdots + \beta_n(a_na_j\sigma^2) = a_0a_j\sigma^2 + \beta_j'\delta_j^2, \qquad j = 1, 2, \cdots n.$$

The proof follows at once using Lemma 3.2 to the equation (4.15).

The following corollary can be easily deduced.

COROLLARY 4.1. *Let the observable random variables $x_j (j = 0, 1, \cdots, n)$ have the linear structural set up $x_j = a_j\xi + \eta_j$, where the a_j's are a set of non-zero constants and the ξ and η_j's are mutually independent proper random variables each having a finite expectation. Then the necessary and sufficient condition for the multiple regression of x_0 on $x_1, x_2, \cdots, x_n$ to be linear (when $n \geqq 2$) is that ξ and $\eta_1, \eta_2, \cdots, \eta_n$ are normally distributed.*

This corollary has been proved earlier independently by the author [5] and Ferguson [3].

5. A theorem in general linear structure. We shall now consider a theorem on characterisation connected with the general linear structural set up already defined by the equation (2.1) in Section 2. In this direction, Ferguson [3] has obtained some necessary and sufficient conditions for the multiple regression of x_0 on $x_1, x_2, \cdots, x_n$ to be linear irrespective of the values of the constants a_{ij}. In the case of the higher dimensional structure, no result has yet been obtained, assuming the regression to be linear for just one set of values of the constants a_{ij}. We shall now show that it is possible to obtain some result for the case of the general linear structural relation for only one set of the values of the constants a_{ij} (with some restrictions upon their values) under the additional assumption that the conditional distribution of x_0 given $x_1, x_2, \cdots, x_n$ is homoscedastic and all the random variables concerned have finite variances.

We are now in a position to prove the following theorem:

THEOREM 5.1. *In the general linear structural set up* (2.1) *and under the Assumptions* 1, 2 *and* 3, *if the constants a_{ij}'s are subject to the following restrictions*

(i) *the vector $\alpha_j = (a_{1j}, a_{2j}, \cdots, a_{nj})$ has at least one non-zero element for each $j = 1, 2, \cdots, p$,*

(ii) *the matrix $(A\Sigma A' + \Delta)$ is non-singular, that is the determinant*

$$|A\Sigma A' + \Delta| \neq 0,$$

(iii) *each of the elements of the vectors* $\alpha_0\Sigma A'(A\Sigma A' + \Delta)^{-1}$ *and*

$$\alpha_0[I - \Sigma A'(A\Sigma A' + \Delta)^{-1}A]$$

is different from zero,
then the necessary and sufficient condition for

$$\begin{cases} E(x_0|x_1, x_2, \cdots, x_n) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n, \\ V(x_0|x_1, x_2, \cdots, x_n) = \sigma_0^2, \end{cases}$$

is that each of $\xi_1, \xi_2, \cdots, \xi_p$ *and each of the proper random variables amongst* $\eta_1, \eta_2, \cdots, \eta_n$ *are normally distributed.*

PROOF.

Necessity. Let $\varphi(t_0, t_1, \cdots, t_n)$ denote the characteristic function of the distribution of $(x_0, x_1, \cdots, x_n)$; $\Phi_j(t)$ that for the distribution of $\xi_j (j = 1, 2, \cdots, p)$ and $\varphi_k(t)$ that for the distribution of $\eta_k (k = 0, 1, 2, \cdots, n)$.

Then it is easy to obtain

$$\begin{aligned} (5.1) \qquad \varphi(t_0, t_1, \cdots, t_n) &= E[\exp(i\textstyle\sum_{k=0}^{n} t_k x_k)] \\ &= \textstyle\prod_{j=1}^{p} \Phi_j(\sum_{k=0}^{n} a_{kj}t_k) \prod_{k=0}^{n} \varphi_k(t_k). \end{aligned}$$

Now under the assumptions of the theorem and applying the equation (3.2) in Lemma 3.1 to (5.1) above, we get after some laborious algebraic computations, proceeding in the same way as in Section 4,

$$(5.2) \qquad \textstyle\sum_{j=1}^{p} (a_{0j} - \sum_{k=1}^{n} \beta_k a_{kj})\Theta_j'(\sum_{k=1}^{n} a_{kj}t_k) = \sum_{k=1}^{n} \beta_k \theta_k'(t_k),$$

$$(5.3) \qquad \textstyle\sum_{j=1}^{p} \{a_{0j}^2 - (\sum_{k=1}^{n} \beta_k a_{kj})^2\}\Theta_j''(\sum_{k=1}^{n} a_{kj}t_k) = -(\sigma_0^2 - \delta_0^2) + \sum_{k=1}^{n} \beta_k^2 \theta_k''(t_k),$$

holding for all real $t_1, t_2, \cdots, t_n$ in a suitably chosen neighbourhood of the origin, where

$$\Theta_j(t) = \ln \Phi_j(t), \qquad j = 1, 2, \cdots p;$$

$$\theta_k(t) = \ln \varphi_k(t), \qquad k = 1, 2, \cdots n.$$

Under the assumption that each of the random variables concerned has a finite second moment, we may again differentiate both sides of the equation (5.2) with respect to $t_l (l = 1, 2, \cdots, n)$ and thus obtain the set of equations

$$(5.4) \qquad \textstyle\sum_{j=1}^{p} a_{lj}(a_{0j} - \sum_{k=1}^{n} \beta_k a_{kj})\Theta_j''(\sum_{k=1}^{n} a_{kj}t_k) = \beta_l \theta_l''(t_l), \quad l = 1, 2, \cdots, n.$$

Next multiplying both sides of the equation (5.4) by β_l and adding for all $l = 1, 2, \cdots, n$, we get

$$(5.5) \qquad \textstyle\sum_{l=1}^{n} \sum_{j=1}^{p} \beta_l a_{lj}(a_{0j} - \sum_{k=1}^{n} \beta_k a_{kj})\Theta_j''(\sum_{k=1}^{n} a_{kj}t_k) = \sum_{l=1}^{n} \beta_l^2 \theta_l''(t_l).$$

Now using the equation (5.3), we get a simplification of the following expression

$$(5.6)\quad \sum_{j=1}^{p}(a_{0j}-\sum_{k=1}^{n}\beta_k a_{kj})^2\Theta_j''(\sum_{k=1}^{n}a_{kj}t_k)+\sum_{k=1}^{n}\beta_k^2\theta_k''(t_k)$$
$$=\sum_{j=1}^{p}\{a_{0j}^2-2a_{0j}(\sum_{k=1}^{n}\beta_k a_{kj})+(\sum_{k=1}^{n}\beta_k a_{kj})^2\}\Theta_j''(\sum_{k=1}^{n}a_{kj}t_k)$$
$$+\sum_{k=1}^{n}\beta_k^2\theta_k''(t_k)$$
$$=-(\sigma_0^2-\delta_0^2)$$
$$-2\sum_{j=1}^{p}\{(\sum_{k=1}^{n}\beta_k a_{kj})(a_{0j}-\sum_{k=1}^{n}\beta_k a_{kj})\}\Theta_j''(\sum_{k=1}^{n}a_{kj}t_k)$$
$$+2\sum_{k=1}^{n}\beta_k^2\theta_k''(t_k).$$

Finally using the equation (5.5) to the right-hand side of (5.6), we obtain

$$(5.7)\quad \sum_{j=1}^{p}(a_{0j}-\sum_{k=1}^{n}\beta_k a_{kj})^2\Theta_j''(\sum_{k=1}^{n}a_{kj}t_k)+\sum_{k=1}^{n}\beta_k^2\theta_k''(t_k)=-(\sigma_0^2-\delta_0^2).$$

Since it is given that the matrix $(A\Sigma A'+\Delta)$ is non-singular, it can be easily shown under the condition $E(x_0|x_1, x_2, \cdots, x_n)=\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n$, β_k is given by the kth element of the vector $a_0\Sigma A'(A\Sigma A'+\Delta)^{-1}$ for $k=1, 2, \cdots, n$. Similarly $a_{0j}-\sum_{k=1}^{n}\beta_k a_{kj}$ is given by the jth element of the vector

$$a_0[I-\Sigma A'(A\Sigma A'+\Delta)^{-1}A] \quad \text{for } j=1, 2, \cdots, p.$$

Thus under the given restrictions on a_{ij}'s, it follows that $a_{0j}-\sum_{k=1}^{n}\beta_k a_{kj}\neq 0$ for all $j=1, 2, \cdots, p$ and similarly $\beta_k\neq 0$ for all $k=1, 2, \cdots, n$. Then the proof of the necessity part follows at once, using Linnik's result (Lemma 3.3) to the equation (5.7).

The proof that the condition is sufficient follows easily from Cramér ([2], pp. 314–315).

REFERENCES

[1] H. V. Allen, "A theorem concerning the linearity of regression," *Stat. Res. Memoirs.*, Vol. 2 (1938), pp. 60–68.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] T. Ferguson, "On the existence of linear regression in linear structural relations," *University of California Publications in Statistics*, Vol. 2 (1955), pp. 143–166.

[4] E. Fix, "Distributions which lead to linear regressions," *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1949, pp. 79–91.

[5] R. G. Laha, "On Characterizations of Probability Distributions and Statistics," Ph.D. thesis submitted to Calcutta University, 1955.

[6] R. G. Laha, "On a characterization of stable law with finite expectation," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 187–195.

[7] R. G. Laha, "On a characterization of the normal distribution from properties of suitable linear statistics," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 126–139.

[8] Yu, V. Linnik, "On an analytical extension of Cramer's theorem on the normal law," *Reports of the Leningrad and Moscow University Probability Seminars*, 1954.

[9] C. R. Rao, M.A. thesis submitted to Calcutta University, 1943.

[10] C. R. Rao, "Note on a problem of Ragnar Frisch," *Econometrica*, Vol. 15 (1947), pp. 245–249.

[11] C. R. Rao, "A correction to 'Note on a problem of Ragnar Frisch,' " *Econometrica*, Vol. 17 (1949), p. 212.

[12] C. Rothschild and E. Mourier, "Sur les lois de probabilité à regression linéaire et écart type lié constant," *C. R. Acad Sci, Paris*, Vol. 225 (1947), pp. 1117–1119.

[13] A. A. Zinger and Yu. U. Linnik, "On an analytical extension of a theorem of Cramér and its applications (Russian)", *Vestnik Leningrad Univ.*, Vol. 10 (1955), pp. 51–56.

RANDOM ORTHOGONAL TRANSFORMATIONS AND THEIR USE IN SOME CLASSICAL DISTRIBUTION PROBLEMS IN MULTIVARIATE ANALYSIS[1]

By Robert A. Wijsman

Department of Statistics, University of California, Berkeley

0. Summary. Orthogonal matrices having elements depending on certain random vectors provide a useful tool in various distribution problems in multivariate analysis. The method is applied to the derivation of the distributions of Hotelling's T^2 and Wilks' generalized variance, the Bartlett decomposition, and the Wishart distribution.

1. Introduction. The purpose of this paper is to demonstrate a method for treating some distribution problems in multivariate normal analysis, and to apply this method to the derivation of the Wishart distribution, the Bartlett decomposition, and the distributions of Hotelling's T^2 and Wilks' generalized variance. A large number of different derivations of these statistics exist in the literature ([1], [2], [4], [6] to [23]), and which one is preferable is a matter of taste. The motivation for presenting yet another derivation of well-known results is that it is believed that the method presented here leads to the results faster than existing derivations, without the necessity of extensive preparation, and almost without computations. A further advantage of the method is that it leads immediately to a representation of the statistics mentioned in terms of combinations of independent normal variables. More specifically, apart from constant factors, Hotelling's T^2 is obtained as an F variable, Wilks' generalized variance as a product of independent χ^2 variables, while the Wishart distribution is simply related to the joint distribution of independent normal and χ^2 variables (Bartlett decomposition [3]). These are known facts, stated explicitly by some authors ([2], [6], [16], [18]), but clearly demonstrated by only few derivations in the literature: Elfving [6] and Ogawa [16] obtain the Bartlett decomposition; Anderson [2] and Elfving [6] obtain the generalized variance essentially as a product of χ^2 variables; while Anderson [2] obtains T^2 essentially as an F variable in a rather indirect way, by relating it to a multiple correlation coefficient. The method presented in this paper will lead to the results in a simple direct, and unified way.

It can be expected that orthogonal transformations provide at least as powerful a tool in the multivariate case as in the univariate case. Indeed, an example can be found in the work of James [12]. The method followed in the present paper will also lean very heavily on orthogonal transformations of random variables. In order to utilize this tool to the utmost, most of the useful transforma-

Received June 18, 1956.

[1] This investigation was supported (in part) by a research grant (RG-3666) from the Institutes of Health, U. S. Public Health Service.

tions will be performed with orthogonal matrices, the elements of which depend on a random vector. This idea is not new, but is often couched in geometrical language.[2] In this respect the treatment in this paper will have something in common with that of Elfving [6] and Ogawa [16]. However, the method followed here does not seem to have appeared in the literature in the same form.

2. Methods and main results. *Notation*: Boldface symbols denote matrices and column vectors, prime denotes transposition, $\mathbf{0}$ is a zero vector, $\mathbf{I}_n$ an $n \times n$ identity matrix, $\mathbf{\Omega}$ an orthogonal matrix. If $\mathbf{A}$ is a square matrix, then $|\mathbf{A}|$ denotes the absolute value of its determinant, and tr $\mathbf{A}$ its trace. $\mathfrak{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the distribution of a normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\mathfrak{N}(0, 1)$ the distribution of a normal variable with zero mean and unit variance. A χ^2 variable with n degrees of freedom is denoted by χ^2_n, an F variable with n_1 and n_2 degrees of freedom by F_{n_1,n_2}. Of a $k \times n$ matrix of variables u_{ir}, the i-th row is denoted by $\mathbf{U}'_i$, the r-th column by $\mathbf{U}_{(r)}$.

The dominant method used throughout this paper is transformation by an orthogonal matrix, the elements of which depend on a random vector. The usefulness of this method depends on the following lemma.

LEMMA 1. *Let* $\mathbf{X}$ *be a random vector with components* $x_1 \cdots x_n$, *and let* $\mathbf{\Omega}(\mathbf{Z})$ *be a random* $n \times n$ *orthogonal matrix whose elements depend in a measurable way on a random vector* $\mathbf{Z}$ *which is independent of* $\mathbf{X}$. *Let* $\mathbf{Y} = \mathbf{\Omega X}$; *then if* $\mathbf{X}$ *is* $\mathfrak{N}(\mathbf{0}, \mathbf{I}_n)$, $\mathbf{Y}$ *is also* $\mathfrak{N}(\mathbf{0}, \mathbf{I}_n)$ *and independent of* $\mathbf{Z}$.

The proof of Lemma 1 follows immediately from the fact that the conditional distribution of $\mathbf{Y}$, given $\mathbf{Z}$, is $\mathfrak{N}(\mathbf{0}, \mathbf{I}_n)$ and is therefore independent of $\mathbf{Z}$.

The lemma will usually be applied in cases where $\mathbf{X}$ and $\mathbf{Z}$ have the same number of components, and $\mathbf{\Omega}(\mathbf{Z})$ is defined in such a way that $\mathbf{\Omega Z}$ has all but its last component equal to zero. In Appendix 1 it will be shown that $\mathbf{\Omega}$ can be uniquely defined in a measurable way.

Throughout this paper we have to consider random matrices, the elements of which are independent $\mathfrak{N}(0, 1)$ variables. A $k \times n$ matrix of independent $\mathfrak{N}(0, 1)$ variables x_{ir} $(i = 1 \cdots k, r = 1 \cdots n)$, will be denoted by $\mathbf{M}^x_{kn}$. We shall assume $k \leqq n$. The i-th row of $\mathbf{M}^x_{kn}$ will be denoted by $\mathbf{X}'_i$ $(i = 1 \cdots k)$. With $\mathbf{M}^x_{kn}$ we form the symmetric matrix $\mathbf{A}^x_{kn}$

$$\mathbf{A}^x_{kn} = \mathbf{M}^x_{kn} (\mathbf{M}^x_{kn})', \tag{1}$$

whose ij-th element is $\mathbf{X}'_i\mathbf{X}_j$ $(i, j = 1 \cdots k)$.

Consider the transformation

$$\mathbf{X}'_i\mathbf{\Omega} = \mathbf{Z}'_i \qquad (i = 1 \cdots k),$$

in which the orthogonal matrix $\mathbf{\Omega}$ depends on $\mathbf{X}_1$ in such a way as to reduce the first $n - 1$ components of $\mathbf{Z}_1$ to 0. The last component of any $\mathbf{Z}_i$—that is, z_{in}—

[2] In a course at Stanford University, Dr. Charles M. Stein uses this idea in the derivation of Hotelling's T^2 distribution (private communication).

will, in the following, be denoted for short by z_i $(i = 1 \cdots k)$. For z_1 we have

$$z_1^2 = \mathbf{X}_1'\mathbf{X}_1, \tag{2}$$

and z_1^2 is clearly a χ_n^2 variable. Inserting the identity matrix $\mathbf{I}_n = \mathbf{\Omega}\mathbf{\Omega}'$ between the two factors on the right-hand side of (1) we obtain

$$\mathbf{A}_{kn}^x = \mathbf{M}_{kn}^x \mathbf{\Omega}\mathbf{\Omega}'(\mathbf{M}_{kn}^x)' = \begin{Vmatrix} z_1^2 & z_1 z_2 & \cdots & z_1 z_k \\ z_1 z_2 & \mathbf{Z}_2'\mathbf{Z}_2 & \cdots & \mathbf{Z}_2'\mathbf{Z}_k \\ \vdots & \vdots & & \vdots \\ z_1 z_k & \mathbf{Z}_k'\mathbf{Z}_2 & \cdots & \mathbf{Z}_k'\mathbf{Z}_k \end{Vmatrix}. \tag{3}$$

Let $\mathbf{Y}_i$ $(i = 1 \cdots k-1)$ be the $(n-1)$-component vector obtained from $\mathbf{Z}_{i+1}$ by deleting its last component z_{i+1}:

$$y_{ir} = z_{i+1,r} \qquad (i = 1 \cdots k-1, r = 1 \cdots n-1),$$

so that we have

$$\mathbf{X}_{i+1}'\mathbf{X}_{j+1} = \mathbf{Z}_{i+1}'\mathbf{Z}_{j+1} = \mathbf{Y}_i'\mathbf{Y}_j + z_{i+1}z_{j+1} \qquad (i, j = 1 \cdots k-1). \tag{4}$$

It is now possible to write (3) in the following way:

$$\mathbf{A}_{kn}^x = \begin{Vmatrix} z_1 & & & \\ z_2 & 1 & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \\ z_k & & & 1 \end{Vmatrix} \begin{Vmatrix} 1 & 0 & \cdots & 0 \\ 0 & \mathbf{Y}_1'\mathbf{Y}_1 & \cdots & \mathbf{Y}_1'\mathbf{Y}_{k-1} \\ \vdots & \vdots & & \vdots \\ 0 & \mathbf{Y}_{k-1}'\mathbf{Y}_1 & \cdots & \mathbf{Y}_{k-1}'\mathbf{Y}_{k-1} \end{Vmatrix} \begin{Vmatrix} z_1 & z_2 & \cdots & z_k \\ & 1 & & \\ & & \cdot & \\ & & & \cdot \\ & & & 1 \end{Vmatrix}, \tag{5}$$

in which z_1 is χ_n, $z_2 \cdots z_k$ and the y_{ir} $(i = 1 \cdots k-1, r = 1 \cdots n-1)$ are $\mathfrak{N}(0, 1)$, and all variables are independent. Equation (5) can be written more concisely if we denote by $\mathbf{Z}$ the vector with components $z_2 \cdots z_k$:

$$\mathbf{A}_{kn}^x = \begin{Vmatrix} z_1 & \mathbf{0}' \\ \mathbf{Z} & \mathbf{I}_{k-1} \end{Vmatrix} \begin{Vmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{A}_{k-1,n-1}^y \end{Vmatrix} \begin{Vmatrix} z_1 & \mathbf{Z}' \\ \mathbf{0} & \mathbf{I}_{k-1} \end{Vmatrix}. \tag{6}$$

Almost everything will follow from (5) or (6), which is essentially the first step in the Bartlett decomposition. Taking determinants in (6) we get

$$|\mathbf{A}_{kn}^x| = z_1^2 |\mathbf{A}_{k-1,n-1}^y|, \tag{7}$$

in which the two factors on the right-hand side are independent, and z_1^2 has a χ_n^2 distribution. Upon repeated application of (7) we get the following result:

LEMMA 2. *The distribution of* $|\mathbf{A}_{kn}^x|$ *is the distribution of the product of* k *independent* χ^2 *variables with* $n, n-1, \cdots, n-k+1$ *degrees of freedom, respectively.*

Equation (7) holds for any k. Writing (7) with k replaced by $k - h$ ($h \leqq k - 1$) and forming ratios, we have

$$\frac{|\mathbf{A}^x_{kn}|}{|\mathbf{A}^x_{k-h,n}|} = \frac{|\mathbf{A}^y_{k-1,n-1}|}{|\mathbf{A}^y_{k-h-1,n-1}|}, \tag{8}$$

in which we set by convention $\mathbf{A}^x_{on} = 1$. Upon iteration of (8) and use of Lemma 2 we get the following result.

LEMMA 3. *The ratio* $|\mathbf{A}^x_{kn}|\,|\mathbf{A}^x_{k-h,n}|^{-1}$, *for* $h \leqq k - 1$, *is distributed like the product of* h *independent* χ^2 *variables with* $n - k + h$, $n - k + h - 1, \cdots, n - k + 1$ *degrees of freedom, respectively.*

COROLLARY. *The ratio* $|\mathbf{A}^x_{kn}|\,|\mathbf{A}^x_{k-1,n}|^{-1}$ *is a* χ^2_{n-k+1} *variable.*

The results so far obtained are sufficient to derive the distributions of T^2 and the generalized variance. For the Wishart distribution, however, it is necessary to consider the joint distribution of the $\frac{1}{2}k(k+1)$ distinct elements $a_{ij} = \mathbf{X}_i'\mathbf{X}_j$ ($i = 1 \cdots k, j = i \cdots k$) of $\mathbf{A}^x_{kn}$. The decomposition (5) expresses the a_{ij} as functions of the χ^2_n variable z_1^2, the $k - 1$ $\mathfrak{N}(0, 1)$ variables $z_2 \cdots z_k$, and the $\frac{1}{2}k(k-1)$ variables $b_{ij} = \mathbf{Y}_i'\mathbf{Y}_j$ ($i = 1 \cdots k - 1, j = i \cdots k - 1$). If the decomposition (5) is continued, then the a_{ij} are expressed as functions of $\frac{1}{2}k(k-1)$ $\mathfrak{N}(0, 1)$ variables and k χ^2 variables with $n, n-1, \cdots, n-k+1$ degrees of freedom, respectively, all variables being independent.[3] The joint distribution of these variables, together with the Jacobian of the transformation, will produce the joint distribution of the a_{ij}.

According to (3), the a_{ij} are first expressed as functions of new variables z_1^2, z_i, and $\mathbf{Z}_i'\mathbf{Z}_j$ ($i = 2 \cdots k, j = i \cdots k$). Subsequently a new set of variables $b_{ij} = \mathbf{Y}_i'\mathbf{Y}_j$ ($i = 1 \cdots k - 1, j = i \cdots k - 1$) is introduced, connected with the $\mathbf{Z}_i'\mathbf{Z}_j$ through (4). The first transformation yields a Jacobian z_1^{k-1}, the second yields unity. Hence, the Jacobian of the transformation from the $\frac{1}{2}k(k+1)$ variables a_{ij} ($i = 1 \cdots k, j = i \cdots k$) to the $\frac{1}{2}k(k+1)$ variables $z_1^2, z_2 \cdots z_k$, and b_{ij} ($i = 1 \cdots k - 1, j = i \cdots k - 1$) is z_1^{k-1}. Let the density of the a_{ij} ($i = 1 \cdots k, j = i \cdots k$) be denoted by $p(\mathbf{A}^x_{kn})$, and the density of the b_{ij} ($i = 1 \cdots k-1, j = i \cdots k-1$) by $p(\mathbf{A}^y_{k-1,n-1})$. The joint density of the $\mathfrak{N}(0, 1)$ variables $z_2 \cdots z_k$, the χ^2_n variable z_1^2, and the variables b_{ij} is given by

$$(2\pi)^{-(1/2)(k-1)} \exp\left[-\tfrac{1}{2}\sum_{i=2}^{k} z_i^2\right] 2^{-(1/2)n/2}\Gamma^{-1}\left(\frac{n}{2}\right) z_1^{n-2} \exp\left[-\tfrac{1}{2}z_1^2\right] p(\mathbf{A}^y_{k-1,n-1}).$$

Taking the Jacobian z_1^{k-1} into account, we have:

$$p(\mathbf{A}^x_{kn}) = c_{kn} z_1^{n-k-1} \exp\left[-\tfrac{1}{2}\left(z_1^2 + \sum_{i=2}^{k} z_i^2\right)\right] p(\mathbf{A}^y_{k-1,n-1}), \tag{9}$$

[3] If the decomposition (5) is continued, the right hand side can be written as the product of a triangular matrix and its transposed, the elements in the triangular matrix being independent normal and χ^2 variables. The decomposition in that form was also obtained by Mauldon [15].

with

$$c_{kn}^{-1} = (2\pi)^{(k-1)/2} 2^{n/2} \Gamma\left(\frac{n}{2}\right). \tag{10}$$

In order to write (9) so that it can be iterated immediately, we observe first that $z_1^2 = |\mathbf{A}_{kn}^x| \, |\mathbf{A}_{k-1,n-1}^y|^{-1}$ by (7). Furthermore,

$$z_1^2 + \sum_{i=2}^{k} z_i^2 = \sum_{i=1}^{k} \mathbf{X}_i'\mathbf{X}_i - \sum_{i=1}^{k-1} \mathbf{Y}_i'\mathbf{Y}_i = \operatorname{tr} \mathbf{A}_{kn}^x - \operatorname{tr} \mathbf{A}_{k-1,n-1}^y,$$

using (2) and (4). Thus we can write (9) in the following way:

$$p(\mathbf{A}_{kn}^x) \, |\mathbf{A}_{kn}^x|^{-(n-k-1)/2} \exp\left[\tfrac{1}{2} \operatorname{tr} \mathbf{A}_{kn}^x\right]$$
$$= c_{kn} \, p(\mathbf{A}_{k-1,n-1}^y) \, |\mathbf{A}_{k-1,n-1}^y|^{-(n-k-1)/2} \exp\left[\tfrac{1}{2} \operatorname{tr} \mathbf{A}_{k-1,n-1}^y\right],$$

from which follows immediately by iteration

$$p(\mathbf{A}_{kn}^x) \, |\mathbf{A}_{kn}^x|^{-(n-k-1)/2} \exp\left[\tfrac{1}{2} \operatorname{tr} \mathbf{A}_{kn}^x\right] = \prod_{i=0}^{k-1} c_{k-i,n-i} = C_{kn},$$

with C_{kn} given by

$$C_{kn}^{-1} = 2^{(1/2)(kn)} \pi^{(1/4)k(k-1)} \prod_{i=0}^{k-1} \Gamma\left(\frac{n-i}{2}\right). \tag{11}$$

We have then, finally,

$$p(\mathbf{A}_{kn}^x) = C_{kn} \, |\mathbf{A}_{kn}^x|^{(n-k-1)/2} \exp\left[-\tfrac{1}{2} \operatorname{tr} \mathbf{A}_{kn}^x\right], \tag{12}$$

with C_{kn} given by (11).

3. Applications. Let $\mathbf{U}_{(1)} \cdots \mathbf{U}_{(n)}$ be n independent observations on a k-component random vector $\mathbf{U}$, which is $\mathfrak{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $k \leqq n - 1$. The components of $\mathbf{U}_{(r)}$ will be denoted by u_{ir} $(i = 1 \cdots k, r = 1 \cdots n)$. The sample mean is $\bar{\mathbf{U}} = (1/n) \sum_{r=1}^{n} \mathbf{U}_{(r)}$, having components $\bar{u}_1 \cdots \bar{u}_k$. The sample covariance matrix $\mathbf{S}$ has components s_{ij} $(i, j = 1 \cdots k)$ given by

$$s_{ij} = \frac{1}{n-1} \sum_{r=1}^{n} (u_{ir} - \bar{u}_i)(u_{jr} - \bar{u}_j). \tag{13}$$

(a) *Hotelling's* T^2. Hotelling's T^2 is defined as

$$T^2 = n(\bar{\mathbf{U}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{U}} - \boldsymbol{\mu}_0), \tag{14}$$

in which $\boldsymbol{\mu}_0$ is some specified vector. T^2 is not defined on the null set in the sample space on which $\mathbf{S}$ is singular.

First consider the case $\boldsymbol{\mu} = \boldsymbol{\mu}_0$. By making the proper transformations it can be shown[4] that $[1/(n-1)]T^2$ has the same distribution as T_1^2 defined by

[4] See, for example [8]. In order to make this paper self-contained, a proof is given in Appendix 2.

$$T_1^2 = \mathbf{X}'_{(n)}(\mathbf{A}^x_{k,n-1})^{-1}\mathbf{X}_{(n)}\,, \tag{15}$$

in which $\mathbf{X}_{(1)} \cdots \mathbf{X}_{(n)}$ are independent, $\mathfrak{N}(\mathbf{0}, \mathbf{I}_k)$, and $\mathbf{A}^x_{k,n-1}$ is defined by (1). In (15) all the variables are independent, and $\mathfrak{N}(0, 1)$. By subjecting the $\mathbf{X}_{(r)}$ to an orthogonal transformation with matrix $\mathbf{\Omega}$: $\mathbf{Y}_{(r)} = \mathbf{\Omega}\mathbf{X}_{(r)}$ $(r = 1 \cdots n)$, we can write (15) also as

$$T_1^2 = \mathbf{Y}'_{(n)}(\mathbf{A}^y_{k,n-1})^{-1}\mathbf{Y}_{(n)}\,, \tag{16}$$

with $\mathbf{A}^y_{k,n-1} = \mathbf{M}^y_{k,n-1}(\mathbf{M}^y_{k,n-1})'$ and $\mathbf{M}^y_{k,n-1} = \mathbf{\Omega}\mathbf{M}^x_{k,n-1}$. The columns of $\mathbf{M}^x_{k,n-1}$ are $\mathbf{X}_{(1)} \cdots \mathbf{X}_{(n-1)}$. Hence, if $\mathbf{\Omega}$ depends only on $\mathbf{X}_{(n)}$, then, by Lemma 1, the elements of $\mathbf{M}^y_{k,n-1}$ are still independent, $\mathfrak{N}(0, 1)$, and independent of $\mathbf{Y}_{(n)}$. We now choose $\mathbf{\Omega}$ such that the first $k-1$ components of $\mathbf{Y}_{(n)}$ are 0. The k-th component of $\mathbf{Y}_{(n)}$ will be denoted by y. From (16) it follows that T_1^2 equals the product of y^2 and the kk-th element of $(\mathbf{A}^y_{k,n-1})^{-1}$, where it has to be remembered that these factors are independent. Now $y^2 = \mathbf{X}'_{(n)}\mathbf{X}_{(n)}$ is a χ^2_k variable, and the kk-th element of $(\mathbf{A}^y_{k,n-1})^{-1}$ equals $|\mathbf{A}^y_{k-1,n-1}| \,/\, |\mathbf{A}^y_{k,n-1}|$, the reciprocal of which is a χ^2_{n-k} variable by the corollary to Lemma 3. Hence T_1^2 is the ratio of two independent χ^2 variables, with k and $(n-k)$ degrees of freedom, respectively. It follows that $(n-1)^{-1}k^{-1}(n-k)T^2$ is an $F_{k,n-k}$ variable.

If $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, then, in (15), $\mathbf{X}_{(n)}$ no longer has zero mean, with the consequence that y^2 is a noncentral χ^2_k variable. On the other hand, the distribution of $\mathbf{A}^x_{k,n-1}$ is unchanged. It follows then that $(n-1)^{-1}k^{-1}(n-k)T^2$ is a noncentral $F_{k,n-k}$ variable. Its distribution was first derived by Hsu [9].

(b) *Wilks' generalized variance.* Wilks' generalized variance is defined as $|\mathbf{S}|$, the determinant of the sample covariance matrix given by (13). By making the same transformation which led to (15) (see also Appendix 2), we find that

$$(n-1)\,\mathbf{CSC}' = \mathbf{A}^x_{k,n-1}\,, \tag{17}$$

in which $\mathbf{C}$ is a nonsingular matrix transforming $\mathbf{\Sigma}$ to the identity matrix

$$\mathbf{C\Sigma C}' = \mathbf{I}_k\,. \tag{18}$$

Taking determinants in (17) and (18) and using Lemma 2, we have then immediately the result that $(n-1)\,|\mathbf{\Sigma}|^{-1}\,|\mathbf{S}|$ is distributed like the product of k independent χ^2 variables with $n-1, \cdots, n-k$ degrees of freedom, respectively. The density of this distribution can be obtained easily only for $k = 1$ and $k = 2$. For $k \geqq 3$, expressions in terms of infinite series have been given by Kullback [13].

(c) *The Wishart distribution and Bartlett decomposition.* The Wishart distribution is the joint distribution of the $\frac{1}{2}k(k+1)$ distinct elements of the sample covariance matrix $\mathbf{S}$, given by (13). It is more convenient to study $\mathbf{S}_1 = (n-1)\mathbf{S}$. By (17) we have

$$\mathbf{CS}_1\mathbf{C}' = \mathbf{A}^x_{k,n-1}\,. \tag{19}$$

The linear transformation (19) relates the sample covariance matrix to $\mathbf{A}^x_{k,n-1}$, defined by (1). The decomposition (5) or (6) of $\mathbf{A}^x_{k,n-1}$ is essentially the first step

in the Bartlett decomposition, giving rise to a χ^2_{n-1} variable z_1^2 and $k-1$ $\mathfrak{N}(0,1)$ variables $z_2 \cdots z_k$. If the decomposition is continued, then $\mathbf{A}^x_{k,n-1}$, and therefore by (19) also $\mathbf{S}_1$, is related in a simple way to k χ^2 variables and $\frac{1}{2}k(k-1)$ $\mathfrak{N}(0,1)$ variables, all independent, which provides the complete Bartlett decomposition.

The density of the $\frac{1}{2}k(k+1)$ distinct elements of $\mathbf{A}^x_{k,n-1}$ is given by (12), after replacing n by $n-1$. The Jacobian of the transformation (19) is[5]

$$\frac{\partial(\mathbf{A}^x_{k,n-1})}{\partial(\mathbf{S}_1)} = |\mathbf{C}|^{k+1}, \tag{20}$$

which we find equals $|\boldsymbol{\Sigma}|^{-\frac{1}{2}(k+1)}$, using (18). Furthermore, by (18) and (19) we have $|\mathbf{A}^x_{k,n-1}| = |\boldsymbol{\Sigma}|^{-1}|\mathbf{S}_1|$ and $\operatorname{tr}\mathbf{A}^x_{k,n-1} = \operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{S}_1$. Substitution of these expressions into (12) gives the Wishart distribution

$$p(\mathbf{S}_1) = C_{k,n-1}|\boldsymbol{\Sigma}|^{-(n-1)/2}|\mathbf{S}_1|^{(n-k-2)/2}\exp[-\tfrac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{S}_1],$$

in which it has to be remembered that $\mathbf{S}_1 = (n-1)\mathbf{S}$, and $C_{k,n-1}$ is given by (11).

4. Acknowledgments. The writer wishes to thank Dr. Henry Scheffé for bringing the problem to his attention and for helpful suggestions, and Dr. Charles M. Stein for providing some valuable references.

APPENDICES

Appendix 1. Let $\boldsymbol{\Omega}^{(m)} = \boldsymbol{\Omega}^{(m)}(\mathbf{Z})$ be an $n \times n$ orthogonal matrix $(1 \leqq m \leqq n-1)$ depending on an n-component column vector $\mathbf{Z}$. If $\mathbf{Z}$ has components $z_1 \cdots z_n$, then the elements $\omega^{(m)}_{ij}$ of $\boldsymbol{\Omega}^{(m)}$ will be defined as follows: $\omega^{(m)}_{ij} = \delta_{ij}$, for $i, j = 1 \cdots m-1, m+2 \cdots n$; $\omega^{(m)}_{mm} = \omega^{(m)}_{m+1,m+1} = z_{m+1}(z_m^2 + z_{m+1}^2)^{-\frac{1}{2}}$; $\omega^{(m)}_{m+1,m} = -\omega^{(m)}_{m,m+1} = z_m(z_m^2 + z_{m+1}^2)^{-\frac{1}{2}}$; and all other off-diagonal elements vanish. If both z_m and z_{m+1} are equal to 0, then we define $\boldsymbol{\Omega}^{(m)}$ to be $\mathbf{I}_n$. The effect of $\boldsymbol{\Omega}^{(m)}(\mathbf{Z})$ applied to $\mathbf{Z}$ is that all components of $\mathbf{Z}$ remain unchanged, except for the m-th and $(m+1)$-st components, of which orthogonal linear combinations are taken such as to make the m-th component equal to 0 and the $(m+1)$-st equal to $(z_m^2 + z_{m+1}^2)^{\frac{1}{2}}$. If we put $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{(n-1)} \cdots \boldsymbol{\Omega}^{(1)}$, where $\boldsymbol{\Omega}^{(1)} = \boldsymbol{\Omega}^{(1)}(\mathbf{Z})$, $\boldsymbol{\Omega}^{(2)} = \boldsymbol{\Omega}^{(2)}(\boldsymbol{\Omega}^{(1)}\mathbf{Z}), \cdots$, then the first $n-1$ components of $\boldsymbol{\Omega}\mathbf{Z}$ are zero and the n-th is $(\sum_{i=1}^n z_i^2)^{\frac{1}{2}}$. $\boldsymbol{\Omega}$ is clearly measurable.

Appendix 2. Put $\mathbf{V}_{(r)} = \mathbf{C}(\mathbf{U}_{(r)} - \boldsymbol{\mu}_0)$, $r = 1 \cdots n$, where $\mathbf{C}$ is any nonsingular $k \times k$ matrix. Substitution in (13) and (14) yields the result that T^2 retains the expression given by (14), with $(u_{ir} - \mu_{oi})$ everywhere replaced by v_{ir} (T^2 is invariant under nonsingular linear transformations of $\mathbf{U}$). If $\mathbf{C}$ is chosen such as to transform $\boldsymbol{\Sigma}$ to $\mathbf{I}_k$ (see (18)), then the v_{ir} are independent and $\mathfrak{N}(0,1)$. It is possible now to write $T_1^2 = [1/(n-1)]T^2$ as follows:

$$T_1^2 = n\overline{\mathbf{V}}'\{\mathbf{M}^v_{kn}(\mathbf{I}_n - \boldsymbol{\Delta}_n)(\mathbf{M}^v_{kn})'\}^{-1}\overline{\mathbf{V}}, \tag{21}$$

in which $\boldsymbol{\Delta}_n$ is an $n \times n$ matrix of which every element equals $1/n$. Let $\boldsymbol{\Omega}$ be an $n \times n$ orthogonal matrix whose last column has all elements equal to $1/(n)^{1/2}$.

[5] See [1], [5]. For completeness, a proof is also indicated in Appendix 3.

We find that $\Omega'(\mathbf{I}_n - \Delta_n)\Omega = \mathbf{J}_n$, where $\mathbf{J}_n$ is obtained from $\mathbf{I}_n$ by replacing the nn-th element by 0. If the variables x_{ir} are related to the v_{ir} by $\mathbf{M}^x_{kn} = \mathbf{M}^v_{kn}\Omega$, then $\mathbf{X}_{(n)} = (n)^{1/2}\overline{\mathbf{V}}$, and the matrix in braces in (21) is

$$\mathbf{M}^v_{kn}\Omega\Omega'(\mathbf{I}_n - \Delta_n)\Omega\Omega'(\mathbf{M}^v_{kn})' = \mathbf{M}^x_{kn}\mathbf{J}_n(\mathbf{M}^x_{kn})' = \mathbf{M}^x_{k,n-1}(\mathbf{M}^x_{k,n-1})' = \mathbf{A}^x_{k,n-1},$$

which proves (15).

Appendix 3. Since any nonsingular matrix $\mathbf{C}$ can be written as the product of elementary matrices, equation (20) need only be verified for the latter ones. Multiplication of a square matrix $\mathbf{A}$ by an elementary matrix results in either of the following elementary operations on $\mathbf{A}$:

(i) interchange of two rows (columns),

(ii) multiplication of a row (column) by a constant $c \neq 0$,

(iii) subtraction of a row (column) from another row (column). The absolute values of the determinants of the corresponding elementary matrices are 1, $|c|$, 1, respectively. The corresponding Jacobians can easily be checked to be 1, $|c|^{k+1}$, 1, respectively. This completes the proof.

REFERENCES

[1] A. C. Aitken, "On the Wishart distribution in statistics," *Biometrika*, Vol. 36 (1949), pp. 59–62.

[2] T. W. Anderson, "Multivariate statistical analysis," mimeographed lecture notes.

[3] M. S. Bartlett, "On the theory of statistical regression, *Proc. Roy. Soc.* Edinburgh, Vol. 53 (1933), pp. 260–283.

[4] H. Cramér, "*Mathematical Methods of Statistics*," Princeton University Press, 1946.

[5] W. L. Deemer and I. Olkin, "The Jacobians of certain matrix transformations useful in multivariate analysis," based on lectures by P. L. Hsu, *Biometrika*, Vol. 38 (1951), pp. 345–367.

[6] G. Elfving, "A simple method of deducing certain distributions connected with multivariate sampling," *Skand. Aktuarietids.*, Vol. 30 (1947), pp. 56–74.

[7] D. Fog, "The geometrical method in the theory of sampling," *Biometrika*, Vol. 35 (1948), pp. 46–54.

[8] H. Hotelling, "The generalization of Student's ratio," *Ann. Math. Stat.*, Vol. 2 (1931), pp. 360–378.

[9] P. L. Hsu, "Notes on Hotelling's generalized T," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 231–243.

[10] P. L. Hsu, "A new proof of the joint product moment distribution," *Proc. Cambridge Philos. Soc.*, Vol. 25 (1939), pp. 336–338.

[11] A. E. Ingham, "An integral which occurs in statistics," *Proc. Cambridge Philos. Soc.*, Vol. 29 (1933), pp. 271–276.

[12] A. T. James, "Normal multivariate analysis and the orthogonal group," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 40–75.

[13] S. Kullback, "An Application of characteristic functions to the distribution problem of statistics," *Ann. Math. Stat.*, Vol. 5 (1934), pp. 264–307.

[14] P. C. Mahalanobis, R. C. Bose, and S. N. Roy, "Normalization of statistical variates and the use of rectangular coordinates in the theory of sampling distributions," *Sankhyā*, Vol. 3 (1937), pp. 1–40.

[15] J. G. Mauldon, "Pivotal quantities for Wishart's and related distributions, and a paradox in fiducial theory," *J. Roy. Stat. Soc.*, Ser. B., Vol. 17 (1955), pp. 79–85.

[16] J. OGAWA, "On the sampling distributions of classical statistics in multivariate analysis," *Osaka Math. J.*, Vol. 5 (1953), pp. 13–52.

[17] I. OLKIN, "On distribution problems in multivariate analysis," *Inst. of Stat. Mimeo. Series*, Vol. 43 (1951), pp. 1–126.

[18] C. R. RAO, *Advanced Statistical Methods in Biometric Research*," John Wiley and Sons, 1952.

[19] G. RASCH, "A functional equation for Wishart's distribution," *Ann. Math. Stat.*, Vol. 29 (1948), pp. 262–266.

[20] S. S. WILKS, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.

[21] J. WISHART, "The generalized product moment distribution in sampling from a normal multivariate population," *Biometrika* Vol. 20A (1928), pp. 32–52.

[22] J. WISHART, "Proofs of the distribution law of the second order moment statistics, *Biometrika*, Vol. 35 (1948), pp. 55–57.

[23] J. WISHART AND M. S. BARTLETT, "The generalized product moment distribution in a normal system," *Proc. Cambridge Philos. Soc.*, Vol. 29 (1933), pp. 260–270.

ON THE DISTRIBUTION OF RANKS AND OF CERTAIN RANK ORDER STATISTICS[1]

By Meyer Dwass

Northwestern University and Stanford University

1. Introduction. Suppose $X_1, \cdots, X_m$ and $X_{m+1}, \cdots, X_N$ are two independent samples from two possibly different populations, and $R_1, \cdots, R_m$ are the ranks of the first m observations in the combined sample and $R_{m+1}, \cdots, R_N$ the ranks of the remaining observations. In the first part of the paper, various moment generating functions connected with these ranks are derived. Of particular interest may be the moment generating function of the Wilcoxon statistic. The asymptotic distribution of a finite number of ranks is derived as $N \to \infty$. The remainder of the paper studies certain aspects of the distribution theory of rank order statistics of the form $\sum_{i=1}^{m} f_N(R_i/N)$. The Wilcoxon statistic and the Hoeffding c_1-statistic are special cases of such a statistic. Many previous studies have been devoted to showing its asymptotic normality. The main purpose of the last half of this paper is to show that for certain combinations of sample sizes m, n, and parent populations, the limiting distribution is non-normal as $m \to \infty$, $n \to \infty$, and $m/N \to 0$.

2. Generating functions for ranks. Throughout this paper we suppose that $X_1, \cdots, X_m, X_{m+1}, \cdots, X_{m+n}$ are $N = m + n$ independent random variables, the first m identically distributed, each with c.d.f. F_1 and the last n identically distributed, each with c.d.f. F_2. We suppose these c.d.f.'s are continuous. By the random variable R_i, the rank of X_i, we mean the number of X_j's less than or equal to X_i. The main object of this section is to write an expression for a generating function for ranks, and the following notation is intended to be useful toward that end. Let $u_0 = -\infty$, $u_{r+1} = \infty$, and

$$u_0 < u_1 < \cdots < u_r < u_{r+1}.$$

Then we denote

$$G_{i,j+1} = G_{i,j+1}(u_j, u_{j+1}) = F_i(u_{j+1}) - F_i(u_j) \qquad (i = 1, 2; j = 0, \cdots, r).$$

Let $i_1, i_2, \cdots, i_r$ be a permutation of the $r = p + q (p \leqq m, q \leqq n)$ integers $1, 2, \cdots, p, m + 1, \cdots, m + q$, and let $e_{i_1}, \cdots, e_{i_r}$ be defined by

$$e_{i_1} = \begin{cases} 1 & \text{if } i_1 \text{ is one of } 1, \cdots, p, \\ 2 & \text{if } i_1 \text{ is one of } m + 1, \cdots, m + q, \end{cases}$$

with similar definitions for $e_{i_2}, \cdots, e_{i_r}$. If $d_1 < d_2 < \cdots < d_r$ is a set of positive integers, they uniquely determine a set of non-negative integers w_1,

Received June 18, 1926; revised October 22, 1956.

[1] Work performed under Office of Naval Research Contract Nonr-225(21).

$w_2, \cdots, w_r$ such that

$$\begin{aligned} d_1 &= w_1 + 1, \\ d_2 &= w_1 + w_2 + 2, \\ &\vdots \\ d_r &= w_1 + w_2 + \cdots + w_r + r. \end{aligned} \tag{1}$$

Conversely, (1) determines for non-negative integers $w_1, \cdots, w_r$ a corresponding set $d_1 < \cdots < d_r$.

We first want to evaluate

$$P\{R_{i_1} = d_1, \cdots, R_{i_r} = d_r; \quad d_1 < \cdots < d_r \leqq N\}$$

for positive integers d_i. By an elementary computation we can write this probability as

$$\sum \frac{(m-p)!}{s_1! \cdots s_{r+1}!} \frac{(n-q)!}{t_1! \cdots t_{r+1}!} \cdot \int_{u_1 < \cdots < u_r} G_{11}^{s_1} \cdots G_{1,r+1}^{s_{r+1}} \cdot G_{21}^{t_1} \cdots G_{2,r+1}^{t_{r+1}} \, dF_{e_{i_1}}(u_1) \cdots dF_{e_{i_r}}(u_r). \tag{2}$$

where, to facilitate printing, e_{i_1} is written as e_{i1} when it occurs as a subscript etc., and where $\sum$ is summation over all non-negative integers s_i, t_i such that

$$\begin{aligned} s_1 + \cdots + s_{r+1} &= m - p, \\ t_1 + \cdots + t_{r+1} &= n - q, \\ s_1 + t_1 &= w_1, \\ s_2 + t_2 &= w_2, \\ &\vdots \\ s_r + t_r &= w_r, \end{aligned}$$

the w_i being determined by the d_i as in (1). Next we recognize that (2) equals the coefficients of $T_{i_1}^{w_1} \cdots T_{i_r}^{w_r}$ in

$$\int_{u_1 < \cdots < u_r} (T_{i_1} G_{11} + \cdots T_{i_r} G_{1,r} + G_{1,r+1})^{m-p} \cdot (T_{i_1} G_{21} + \cdots + T_{1_r} G_{2,r} + G_{2,r+1})^{n-q} \, dF_{e_{i1}}(u_1) \cdots dF_{e_{i_r}}(u_r). \tag{3}$$

Since we can write

$$T_{i_1}^{w_1} T_{i_2}^{w_2} \cdots T_{i_r}^{w_r} = \left(\frac{T_{i_1}}{T_{i_2}}\right)^{d_1} \left(\frac{T_{i_2}}{T_{i_3}}\right)^{d_2} \cdots \left(\frac{T_{i_{r-1}}}{T_{i_r}}\right)^{d_{r-1}} (T_{i_r})^{d_r} \cdot \frac{1}{T_{i_1} \cdots T_{i_r}}$$

this suggests that we make the relabelling

$$\frac{T_{i_1}}{T_{i_2}} = \theta_{i_1}, \cdots, \frac{T_{i_{r-1}}}{T_{i_r}} = \theta_{i_{r-1}}, \qquad T_{i_r} = \theta_{i_r} \tag{4}$$

or

$$\begin{aligned} T_{i_1} &= \theta_{i_1} \cdots \theta_{i_r}, \\ T_{i_2} &= \theta_{i_2} \cdots \theta_{i_r}, \\ &\vdots \\ T_{i_r} &= \theta_{i_r}. \end{aligned} \tag{5}$$

Substituting from (5) into (3) and denoting the resulting function of $\theta_1, \cdots, \theta_r$ by $\varphi(i_1, \cdots, i_r)$, we have

$$\sum P(R_{i_1} = d_1, \cdots, R_{i_r} = d_r)\theta_{i_1}^{d_1} \cdots \theta_{i_r}^{d_r} = \theta_{i_1}\theta_{i_2}^2 \cdots \theta_{i_r}^r \varphi(i_1, \cdots, i_r),$$

where $\sum$ is summation over all integers d_i such that $1 \leqq d_1 < d_2 < \cdots < d_r \leqq N$. We can now state the following.

THEOREM 1. *The generating function of* $R_1, \cdots, R_p, R_{m+1}, \cdots, R_{m+q}$ equals

$$\sum P(R_1 = d_1, \cdots, R_{m+q} = d_r)\theta_1^{d_1} \cdots \theta_r^{d_r} = \sum{}' \theta_{i_1}\theta_{i_2}^2 \cdots \theta_{i_r}^r \varphi(i_1, \cdots, i_r), \tag{6}$$

where $\sum$ *is over all possible integers* $d_1, \cdots, d_r$ *between* 1 *and* N *(no two equals to each other) and* $\sum'$ *is over all permutations* $i_1, \cdots, i_r$ *of the integers* $1, \cdots, p, m+1, \cdots, m+q$.

REMARK. Equality among any two d_i's is equivalent to tied ranks which is excluded with probability one by the assumption of continuity of F_1, F_2.

3. Several special cases. In this section we look at three special cases. They will be referred to again later.

A. *The generating function for a single rank.* To find the generating function for a single rank, say R_1 to be specific, set

$$p = 1, \qquad q = 0, \qquad r = 1, \qquad \theta_1 = \theta$$

in (6). We then obtain that

$$\begin{aligned} E\theta^{R_1} &= \theta \int_{-\infty}^{\infty} (\theta G_{11} + G_{12})^{m-1} (\theta G_{21} + G_{22})^n \, dF_1(u) \\ &= \theta \int_{-\infty}^{\infty} ((\theta - 1) F_1(u) + 1)^{m-1} ((\theta - 1) F_2(u) + 1)^n \, dF_1(u). \end{aligned} \tag{7}$$

B. *The generating function for* $R_1, R_2, \cdots, R_m$. For this case we set

$$p = r = m, \qquad q = 0,$$

in (6) and obtain

$$\begin{aligned} &E\theta_1^{R_1}\theta_2^{R_2} \cdots \theta_m^{R_m} \\ &\quad = \sum \theta_{i_1}\theta_{i_2}^2 \cdots \theta_{i_m}^m \int_{u_1 < \cdots < u_m} [\theta_{i_2} \cdots \theta_{i_m}(\theta_{i_1} - 1) F_2(u_1) \\ &\qquad + \cdots + (\theta_{i_m} - 1) F_2(u_m) + 1]^n \, dF_1(u_1) \cdots dF_1(u_m), \end{aligned}$$

where $\sum$ is over all permutations $i_1, i_2, \cdots, i_m$ of $1, 2, \cdots, m$.

C. *The Wilcoxon statistic.* This statistic is $R_1 + \cdots + R_m$. In case B above, set $\theta_1 = \theta_2 = \cdots = \theta_m = \theta$ and we find that

$$E\theta^{R_1+\cdots+R_m} = m!\theta^{m(m+1)/2} \int_{u_1<\cdots<u_m} [\theta^{m-1}(\theta - 1) F_2(u_1) + \cdots + (\theta - 1) F_2(u_m) + 1]^n \, dF_1(u_1) \cdots dF_1(u_m).$$

4. Limiting distributions involving a fixed number of ranks.

A. *A single rank.* From (7) we have that

$$Ee^{i\theta R/N} = e^{i\theta/N} \int_{-\infty}^{\infty} ((e^{i\theta/N} - 1) F_1(u) + 1)^{m-1} ((e^{i\theta/N} - 1) F_2(u) + 1)^n \, dF_1(u).$$

Suppose $m \to \infty$, $n \to \infty$, $m/N \to \rho$. Since

$$|(e^{i\theta/N} - 1)F_j(u) + 1|^2 = |F_j(u)^2 + (1 - F_j(u))^2 + 2(1 - F_j(u))F_j(u) \cos(\theta/N)| \leqq 1,$$

and since, as $N \to \infty$,

$$((e^{i\theta/N} - 1)F_j(u) + 1)^N \to \exp[i\theta F_j(u)] \qquad j = 1, 2,$$

we can apply the Lebesgue bounded convergence theorem to conclude that

$$E \exp(i\theta R_1/N) \to \int_{-\infty}^{\infty} \exp[i\theta(\rho F_1(u) + (1 - \rho) F_2(u))] \, dF_1(u), \tag{8}$$

as $N \to \infty$. Hence R_1/N is asymptotically distributed as a random variable

$$\rho F_1(X) + (1 - \rho)F_2(X),$$

where X has c.d.f. $F_1(X)$.

REMARKS. (a) Notice that the extremes, $\rho = 0$ and $\rho = 1$ are included in this result.

(b) If we do the above computation for R_j/N, its limiting characteristic function is given by the right side of (8) for $j = 1, \cdots, m$, and by the right side of (8) with dF_1 replaced by dF_2 for $j = m + 1, \cdots, m + n$.

(c) A similar analysis shows that $R_{j_1}/N, \cdots, R_{j_k}/N$ are asymptotically independent as $N \to \infty$ if $j_1 < j_2 < \cdots < j_k$ are fixed indices which do not depend on N.

B. $R_1, \cdots, R_m$. We hold m fixed and let $n \to \infty$. Then by the above remarks,

$$E \exp[i(\theta_1 R_1/N + \cdots + \theta_m R_m/N)] \to \prod_{j=1}^{m} \int_{-\infty}^{\infty} \exp[i\theta_j F_2(u)]$$

as $n \to \infty$. Thus, $R_1/N, \cdots, R_m/N$ are asymptotically independent and each is distributed as a random variable $F_2(X)$, where X has c.d.f. F_1. Let us denote the limiting c.d.f. of R_1/N by $S(t)$. That is, there is a c.d.f. S, such that at any

continuity point t of S,

$$P(R_1/N \leqq t) \to S(t), \tag{9}$$

as $N \to \infty$, for fixed m. Notice that in case F_2 has an inverse F_2^{-1}, then $S(t) = F_1(F_2^{-1}(t))$.

5. Limiting distributions of $S = \sum_{i=1}^{m} f_N(R_i/N)$.

In this section we study the asymptotic distribution of rank order statistics of the form

$$S_N = \sum_{i=1}^{m} f_N(R_i/N), \tag{10}$$

where $f_N(i/N)$ is a real number defined for $i = 1, \cdots, N$. We give below a short discussion on why S_N is of interest and on some of the known results regarding its asymptotic distribution.

For convenience, suppose that

$$f_N(1/N) \leqq f_N(2/N) \leqq \cdots \leqq f_N(N/N).$$

Let $H_N(t)$, $(0 < t < \infty)$ be the c.d.f. of the N numbers $f_N(i/N)$. That is, $H_N(t) =$ proportion of $f_N(i/N)$ less than t. Perhaps the most notable example of a statistic of the form (10) is the Wilcoxon statistic, in which case $f_N(t) = t$, $(0 \leqq t \leqq 1)$, for all N. In [3] it was shown that in case F_1, F_2 depend on a single parameter θ and $F_1 = F_2$ when $\theta = 0$, then often a test of $H_0: \theta = 0$ against $H: \theta > 0$ based on (10) for suitably chosen f_N is a locally most powerful rank order test (local in the sense that θ is close to zero). Studies relevant to the asymptotic normality[2] of (10) can be found in [1], [2], [3], [4], [6], [8]. In particular, it may be worth while to mention some specific conditions which insure the asymptotic normality of S_n. In each case we assume that $m/N \to \rho$ as $N \to \infty$ and $0 < \rho < 1$.

(1) $F_1 = F_2$. $f_N(i/N) = EZ_{Ni}^k$ for some positive integer k, where $Z_{N1} \leqq \cdots \leqq Z_{NN}$ are the ordered values of N independent identically distributed random variables [1].

(2) $F_1 = F_2$. The assumption preceding Theorem 2 below holds and the c.d.f. H has its first two moments [3].

(3) $F_1 \neq F_2$ or $F_1 = F_2$. $f_N(t) = f(t) =$ a polynomial in t, which does not depend on N [3].

We shall now construct the examples referred to in the introduction. The main tools are Theorems 2 and 3 which follow. In addition to the basic assumptions made at the beginning of Section 2, we assume also through the remainder of the paper that there is a c.d.f. $H(t)$ such that at every continuity point of

[2] Whenever we refer to the asymptotic distribution we mean a limiting distribution of $(S_N - a_N)/b_N$, as $m \to \infty$, $n \to \infty$ for a proper choice of $\{a_N\}$, $\{b_N\}$. It may be that m depends on n.

$H(t)$,

$$H_N(t) \to H(t), \qquad \text{as } N \to \infty.$$

THEOREM 2. *Let $t_1, \cdots, t_m$ be such that they are continuity points of $H(t)$ and such that $H(t_1), \cdots, H(t_m)$ are continuity points of $S(u)$. Then*

$$P\{f_N(R_1/N) < t_1, \cdots, f_N(R_m/N) < t_m\} \to S(H(t_1)) \cdots S(H(t_m))$$

as $N \to \infty$, provided m is fixed.

PROOF. We can write

$$P\{f_N(R_1/N) < t_1, \cdots, f_N(R_m/N) < t_m\}$$
$$= P\{R_1/N \leqq H_N(t_1), \cdots, R_m/N \leqq H_N(t_m)\}.$$

The result follows from (9) and the remarks preceding it.

COROLLARY. *Let $\varphi(u)$ be the characteristic function of a random variable with* c.d.f. $R(t) = S(H(t))$. *Then*

$$E \exp i(u_1 f_N(R_1/N) + \cdots + u_m f_N(R_m/N)) \to \varphi(u_1) \cdots \varphi(u_m)$$

as $N \to \infty$, m fixed.

LEMMA 1. *Let $X_{11}, X_{12}, \cdots; X_{21}, X_{22}, \cdots$, be two infinite sequences of random variables, and let the random variable*

$$t_{m,n} = t_{m,n}(X_{11}, \cdots, X_{1m}; X_{21}, \cdots, X_{2n})$$

be a function of the $m + n$ random variables in parentheses. Let $\varphi_{m,n}(u) = E \exp(iut_{m,n})$. Suppose

(a) *There is a characteristic function φ such that for every positive integer m and every real u,*

$$\varphi_{m,n}(u) \to [\varphi(u)]^m$$

as $n \to \infty$ and m is fixed.

(b) *There are norming constants a_m, b_m and a characteristic function Ψ, such that for every real u*

$$\exp(-a_m u/b_m)[\varphi(u/b_m)]^m \to \Psi(u) \tag{11}$$

as $m \to \infty$. Then there is a sequence of pairs of positive integers $(1, n(1)), (2, n(2)), \cdots, (m, n(m)), \cdots$ such that

$$\exp(-ia_m u/b_m)[\varphi_{m,n}(u/b_m)] \to \Psi(u) \tag{12}$$

as $m \to \infty$, $n \to \infty$, provided $n \geqq n(m)$.

The proof is elementary and we omit it. We point out that (12) says that the distribution of $(t_{m,n} - a_m)/b_m$ converges to that distribution whose characteristic function is Ψ.

LEMMA 2. Let $R(t) = S(H(t))$ be as defined above. Suppose $0 < F_2(t) < 1$ if and only if $0 < F_1(t) < 1$. Suppose also that $t = F_2^{-1}(u)$, the inverse of $F_2(t)$,

is defined for all $0 < u < 1$. Then

(a) if $F_1 = F_2$, $R(t) = H(t)$,

(b) if $F_2 = H$, $R(t) = F_1(t)$.

PROOF. The proof follows from the fact that $S(u) = F_1(F_2^{-1}(u))$.

THEOREM 3. *Suppose that if $Y_1, Y_2, \cdots$ is an infinite sequence of independent identically distributed random variables, each with* c.d.f. $R(t) = S(H(t))$, *then then there are norming constants* $\{a_m\}, \{b_m\}$, *such that the* c.d.f. *of* $(\sum_1^m Y_i - a_m)/b_m$ *converges to a* c.d.f. $L(t)$, *as* $m \to \infty$. *Then there is a sequence* $(1, n(1)), \cdots, (m, n(m)), \cdots$, *such that the* c.d.f. of

$$\frac{\sum_{i=1}^{m} f_N(R_i/N) - a_m}{b_m} \tag{13}$$

converges to $L(t)$ *as* $m \to \infty$, $n \to \infty$, *provided that* $n \geqq n(m)$.

PROOF. Let $t_{m,n} = \sum_{i=1}^m f_N(R_i/N)$. This statistic satisfies Condition (a) of Lemma 1 by the corollary to Theorem 2. (Let $u_1 = \cdots = u_m = u$ in that corollary.) Condition (b) holds by assumption and this completes the proof.

REMARKS. (a) An unsatisfactory feature of this result is that it tells nothing about the relative orders of m and n. It is clear that we can find sequences $\{m_i\}$, $\{n_i\}$, such that if $m = m_i$, $n = n_i$, then the asymptotic distribution of the proposition holds and

$$\lim_i m_i/(m_i + n_i) = 0. \tag{14}$$

Though our methods here are not sensitive enough to yield this information, the sense of the derivation is such to make reasonable the conjecture that this asymptotic distribution holds for *all* sample size sequences for which (14) holds.

(b) By Lemma 2, if $F_2 = H$ then $R(t) = F_1(t)$. By the proper choice of F_1 we can determine the limiting distribution L to be *any stable* distribution. For example, suppose (10) is the Hoeffding c_1-statistic [7]. That is, $f_N(i/N) = EZ_{Ni}$, where $Z_{N1} \leqq \cdots \leqq Z_{NN}$ are the ordered values of N independent $N(0, 1)$ random variables. According to [5], H is the unit normal c.d.f. Now suppose that the alternative to the usual null hypothesis that $F_1 = F_2$ is that F_2 is the unit normal c.d.f. and that F_1 is the Cauchy c.d.f., centered at θ. Then there are sequences $\{m, n(m)\}$ such that $[\sum_{i=1}^m f_N(R_i/N) - m\theta]/m$ has asymptotically the Cauchy distribution centered at zero. This is so because Lemma 2 (Case (a)) insures that R is the Cauchy c.d.f. and because an average of independent and identically distributed Cauchy variables is distributed like any one of its components.

(c) In case $H(t)$ concentrates all its mass on a bounded interval, then so does $R(t)$ and excluding the one point limiting distribution, the limiting distribution of this theorem must be normal. This will happen if $f_N(t)(0 \leqq t \leqq 1)$ is a polynomial in t which does not depend on N. This is not surprising since for this case [3] shows that if $\lim_{N\to\infty} m/N = \rho$ exists, then S_N is asymptotically normal for all $0 < \rho < 1$. As a matter of fact, these results would seem to imply that one should be able to include the extreme cases $\rho = 0$ and $\rho = 1$. Similarly, if $F_1 = F_2$

and H has its first two moments, then S_N is asymptotically normal. This is also not surprising since for this case [2] shows asymptotic normality for $0 < \rho < 1$.

(d) We can construct further examples of non-normal limiting distributions by supposing $F_1 = F_2$ and by choosing H properly, since by Lemma 2(b), $R = H$. This is presumably of lesser interest than the construction given in Remark (b) above.

REFERENCES

[1] M. Dwass, "On the asymptotic normality of certain rank order statistics," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 303–306.

[2] M. Dwass, "On the asymptotic normality of some statistics used in non-parametric tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 334–339.

[3] M. Dwass, "The large sample power of rank order tests in the two-sample problem," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 352–373.

[4] W. Hoeffding, "A combinatorial central limit theorem," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 558–66.

[5] W. Hoeffding, "On the distribution of the expected values of the order statistics," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 93–100.

[6] G. E. Noether, "On a theorem by Wald and Wolfowitz," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 455–58.

[7] M. E. Terry, "Some rank order tests which are most powerful against specific parametric alternatives," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 346–66.

[8] Wald, A., and J. Wolfowitz, "Statistical tests based on permutations of the observations," *Ann. Math. Stat.* Vol. 15 (1944), pp. 358–72.

ESTIMATING FUTURE FROM PAST IN LIFE TESTING

By John E. Walsh

Military Operations Research Division, Lockheed Aircraft Corp., Burbank, Calif.

1. Summary. Let θ_p represent the unique $100p$ per cent point of a continuous statistical population, while x_r is the rth largest value of a sample of size n from this population ($r = 1, \cdots, n$). This paper considers estimation of θ_p on the basis of $x_{r(1)}, \cdots, x_{r(m)}$, where the $r(i)$ differ by $O(\sqrt{n+1})$ and do not necessarily have values near $(n+1)p$. Also considered is estimation of x_R on the basis of $x_{r(1)}, \cdots, x_{r(m)}$, where the $r(i)$ differ by $O(\sqrt{n+1})$ and do not necessarily have values near R. The results are of a nonparametric nature and based on expected value considerations. These estimation procedures may be useful for life-testing situations where time to failure is the variable and some of the items tested have not yet failed when observation is discontinued. Then θ_p and x_R can be estimated for p and R values which extend a moderate way into the region where sample data is not available. Estimation of the x_R value which would be obtained by continuing to observe the experiment represents a prediction of the future from the past. The results of this paper may be of value in the actuarial, population statistics, operations research, and other fields.

2. Introduction. Let us consider a sampling situation where n items are simultaneously life tested to determine their times to failure. Then the time to failure for the first item which fails is the smallest value for this sample of size n. The value for the second item to fail is the next to smallest sample value; etc. Thus life-testing situations have the property that the r smallest order statistics of a sample are determined in advance of the remaining values of the sample. Moreover, the first r items to fail furnish the r smallest values of the sample of size n, even if some or all of the remaining sample values are never determined. Jacobson called attention to these valuable properties of life-testing situations in [1]. A descriptive outline of the life-testing field is given in [2].

The property that the r smallest order statistics of a lifetesting sample can be obtained without the necessity of determining the remaining sample values can be exploited in many ways. The basis for this exploitation is that substantial time and/or cost can often be saved by stopping a life-testing experiment at some convenient time before all the items have failed. The situation of this type considered here is the estimation of θ_p on the basis of $x_1, \cdots, x_r$ when $(n+1)p > r$—that is, estimation of population percentage points in the region not covered by the available data.

The life-testing property that the r smallest order statistics are determined in advance of the remaining sample values furnishes an opportunity for estimating the future from the past. Suppose that r items have failed up to the present time and it is desired to predict the future time at which the Rth item of this set will

Received June 20, 1956; revised September 14, 1956.

fail ($R > r$). That is, $x_1, \cdots, x_r$ are known and estimation of x_R in an expected value sense is desired. This paper derives such an estimate for the case where r and R do not differ too much.

The purpose of this paper is to derive nonparametric expected value estimates which are approximately valid for nearly all continuous statistical populations of practical interest. These estimates are not intended to be competitive with those which can be obtained on the basis of additional information about the population sampled. Instead, the nonparametric estimates presented are for use when more specialized estimation methods are not warranted.

Whether θ_p or x_R is estimated, the arithmetical problem consists in determining the values of $c_1, \cdots, c_m$ for a linear function of the form

$$\sum_{i=1}^{m} c_i x_{r(i)},$$

where m, $r(1), \cdots, r(m)$ are given integers. The procedure for obtaining the values of the c_i consists in solving m specified linear equations in m unknowns. Although the emphasis of the paper is on life-testing situations, the results derived are valid for more general types of situations than those where only $x_1, \cdots, x_r$ are available and $r < (n+1)p$, R. For estimation of θ_p, knowledge of the values of order statistics $x_{r(1)}, \cdots, x_{r(m)}$ such that $r(i) = r(j) + O(\sqrt{n+1})$ and none of the $r(i)$ differ too much from $(n+1)p$ is sufficient. In estimating x_R, it is sufficient that $x_{r(1)}, \cdots, x_{r(m)}$ are available with $r(i) = r(j) + O(\sqrt{n+1})$ and none of the $r(i)$ differing too much from R. Thus the results of this paper can also be used to estimate the past from the present for life-testing situations where the data for the past was lost or not recorded.

Life-testing situations where population properties are of greater interest than sample properties usually involve inanimate objects such as automobile tires, light bulbs, etc. Often a considerable savings in time and/or expense can be obtained by deliberately stopping a life test of this type when 80 to 90 per cent rather than all of the items have failed. Through use of the method given in this paper, many of the upper population percentage points of interest can be estimated even though the upper 10 per cent to 20 per cent of the data is truncated.

The future mortality occurrences among the now-surviving members of a given set of items can be of interest for some types of life-testing situations. The future mortality among the survivors of a specified group of persons which have already been observed for some time represents a situation of this nature. Estimates of the future mortality among the survivors of such a group of persons can be valuable in actuarial science, population statistics, and other fields. This paper pre sents a rather widely applicable procedure for estimating the first time at which a specified number of additional individuals will have died on the basis of the times to death for the individuals which have already died.

An investigation is made of the variances for the derived estimates. Every estimate considered has an estimate of the form

$$p_i(1 - p_i)/n[f(\theta_{p_i})]^2 + O(n^{-3/2}),$$

where t is a specified number which differs from the $r(i)$ by $O(\sqrt{n+1})$, $p_t = t/(n+1)$, and $f(x)$ is the probability density function (*pdf*) of the statistical population sampled. Thus all the estimates presented are consistent, having standard deviations which are $O(1/\sqrt{n})$.

If all the sample values were available, the corresponding sample percentage point might appear to be the most suitable nonparametric expected value estimate for θ_p. In many cases, however, an estimate for θ_p of the type given in this paper may have a higher efficiency (i.e., smaller variance) than the corresponding sample percentage point. The variance for the sample percentage point corresponding to θ_p is

$$p(1-p)/n[f(\theta_p)]^2 + O(n^{-2}).$$

If n is large and

$$p_t(1-p_t)/f(\theta_{p_t})^2 < p(1-p)/f(\theta_p)^2,$$

the sample percentage point corresponding to θ_p usually has a lower efficiency than an estimate of the type presented here. This inequality is frequently satisfied for unimodal populations where θ_{p_t} is more toward the central part of the probability distribution than θ_p.

In deriving the results, $f(x)$ is assumed to exist, be positive, and of an analytical nature for all x of interest. Such strong restrictions on $f(x)$ are not necessary for the validity of the results presented. However, little generality is gained for practical cases by using weaker restrictions on $f(x)$. There are limitations on the accuracy to which measurements on continuous observations can be made for all applied situations. This data-accuracy limitation indicates that the conditions imposed on $f(x)$ should be acceptable for virtually all practical situations of a continuous type where θ_p is unique for all p.

Section 3 contains a statement of the estimates for θ_p and x_R along with some restrictions on their use. A numerical example of the application of each type of estimate is given in Section 4. Assuming a standard normal population, the approximate properties stated for these estimates are compared with their exact properties. Section 5 contains the derivations and motivation for the material given in Section 3.

3. Statement of estimates. Let us consider an explicit statement of the method for obtaining estimates of θ_p and x_R. The additional notation used is

$$t = \frac{1}{m}\sum_{i=1}^{m} r(i),$$

$$r = \max_{1 \le i \le m} r(i),$$

$$d(i) = t - r(i) = \text{quantity which is } O(\sqrt{n+1}), \qquad (i = 1, \cdots, m);$$

$$d(i) \neq d(k) \text{ if } i \neq k$$

$$p_t = t/(n+1), \qquad q_t = 1 - p_t,$$

$$p_R = R/(n+1), \qquad q_R = 1 - p_R,$$

$$A_j = [\tfrac{1}{2}(p - p_i)]^{j-1}/(j-1)!,$$

$$B_j = \frac{[\tfrac{1}{2}(p_R - p_i)]^{j-1}}{(j-1)!} + \frac{p_R q_R (j-1)(j-2)}{2(n+2)(j-1)!}\,[\tfrac{1}{2}(p_R - p_i)]^{j-2},$$

$$\begin{aligned} C_j[d(i), \delta] = {} & \frac{(-\delta)^{j-1}}{(j-1)!} - \frac{(j-1)(-\delta)^{j-2}\, d(i)}{(n+1)(j-1)!} \\ & + \frac{(j-1)(j-2)(-\delta)^{j-3}}{2(n+1)(n+2)(j-1)!}\,[(n+1)\,p_i q_i + (p_i - q_i)\,d(i) + d(i)^2] \\ & - \frac{(j-1)(j-2)(j-3)(-\delta)^{j-4}\, d(i)}{2(n+1)^2(n+2)(j-1)!}\,[(n+1)\,p_i q_i + \tfrac{1}{3}\,d(i)^2], \end{aligned}$$

$$C(j, i) = C_j[d(i), \tfrac{1}{2}(p - p_i)],$$

$$C'(j, i) = C_j[d(i), \tfrac{1}{2}(p_R - p_i)],$$

where $j = 1, \cdots, m$.

For specified n, m, $r(1), \cdots, r(m)$, p, R, the estimates considered and their principal expected value properties are given by

$$E\left[\sum_{i=1}^{m} a_i x_{r(i)}\right] = \theta_p + O(n^{-3/2}) + O[\,|\,\tfrac{1}{2}(p - p_i)\,|^m],$$

$$E\left[\sum_{i=1}^{m} b_i x_{r(i)}\right] = E(x_R) + O(n^{-3/2}) + O[\,|\,\tfrac{1}{2}(p_R - p_i)\,|^m].$$

If $m \geqq 4$, $O(n^{-3/2})$ is replaced by $O(n^{-2})$ in these expressions. The sets of linear equations used to determine the values of the a_i and the b_i are

$$\sum_{i=1}^{m} a_i C(j, i) = A_j, \qquad \sum_{i=1}^{m} b_i C'(j, i) = B_j, \qquad j = 1, \cdots, m.$$

The values for the a_i and the b_i can be conveniently expressed in the form of determinants. This form is especially useful for small values of m. Explicitly,

$$a_i = \frac{\text{determinant of the } C(u, v), \text{ with } C(j, i) \text{ replaced by } A_j \text{ for } j = 1, \cdots, m}{\text{determinant of the } C(u, v)},$$

$$b_i = \frac{\text{determinant of the } C'(u, v) \text{ with } C'(j, i) \text{ replaced by } B_j \text{ for } j = 1, \cdots, m}{\text{determinant of the } C'(u, v)}.$$

If the determinant of the $C(u, v)$ is zero or near zero, a change in the values for the $r(i)$ may be required to assure that none of the a_i are of too large a magnitude. Usually a change of one value is enough to eliminate this difficulty. The same will be true if the determinant of the $C'(u, v)$ is zero or near zero.

Let us consider determination of a value for m which seems large enough to assure that the unstated higher order expected value terms can be neglected. Accuracy to terms of order $n^{-3/2}$ implies that $m \geqq 3$. The value used is also re-

quired to satisfy the condition

$$m \geqq \begin{cases} \dfrac{\log\,[\max\,(n^{-3/2}, 10^{-4})]}{\log\,|\,\frac{1}{2}(p - p_i)\,|}, & \text{if } \theta_p \text{ estimated,} \\ \dfrac{\log\,[\max\,(n^{-3/2}, 10^{-4})]}{\log\,|\,\frac{1}{2}(p_R - p_i)\,|}, & \text{if } x_R \text{ estimated.} \end{cases}$$

This second condition for determining m is based on the requirement that

$$\max(n^{-3/2}, 10^{-4}) \geqq \begin{cases} |\,\frac{1}{2}(p - p_i)\,|^m, & \text{if } \theta_p \text{ estimated,} \\ |\,\frac{1}{2}(p_R - p_i)\,|^m, & \text{if } x_R \text{ estimated.} \end{cases}$$

These two conditions assure that the expected value error of an estimate is $O[\max\,(n^{-3/2}, 10^{-4})]$. The minimum value of $m = 3$ is acceptable for many of the cases encountered. From a computational viewpoint, the method probably should not be used if the value of m obtained by this procedure exceeds 10.

The convergence rates of the expansions used in obtaining estimates for θ_p depend on $\frac{1}{2}(p - p_i)$, the $d(i)/(n + 1)$, and the properties of the underlying statistical population. In practice, the underlying population properties are usually such that convergence is more rapid for p_i and p near the center of the distribution. On this basis, both $|\,d(i)\,|/(n + 1)p_iq_i$ and $\frac{1}{2}|\,p - p_i\,|/[\min(pq, p_iq_i)]$ should not be too large. The maximum allowable value for these quantities is taken to be $\frac{2}{3}$ for the type of situations considered. This value is not overly small but should be satisfactory for a large majority of the practical applications. Hence, the method given in this paper for estimating θ_p should not be used if either

$$\max_{1 \leq i \leq m} |\,d(i)\,| > \tfrac{2}{3}p_iq_i(n + 1),$$

or

$$|\,p - p_i\,| > \tfrac{4}{3} \min\,(pq, p_iq_i).$$

On a similar basis, the method for estimating x_R should not be used if either

$$\max_{1 \leq i \leq m} |\,d(i)\,| > \tfrac{2}{3}p_iq_i(n + 1),$$

or

$$|\,p_R - p_i\,| > \tfrac{4}{3} \min\,(p_Rq_R\,,\, p_iq_i).$$

Sometimes the inequality involving the $d(i)$ can be changed from unacceptable to acceptable by using a different value for m which allows a decrease in max $|\,d(i)\,|$.

When $x_1\,, \cdots, x_r$ are given and $r > (n + 1)p$, R, a recommended selection for the values of the $r(i)$ in both types of estimates is

$$r(i) = r - (m - i)K, \qquad (i = 1, \cdots, m),$$

where

$$K = \max [1, \text{largest integer contained in } \frac{1}{m}\sqrt{n+1}].$$

The resulting $r(i)$ differ by $O(\sqrt{n+1})$, are equally spaced, and have desirable properties with respect to the expansions used in deriving the estimates.

Every estimate of the two types considered has approximately the same variance. For all estimates derived, the variances are of the form

$$p_t q_t / n[f(\theta_{p_t})]^2 + O(n^{-3/2}).$$

Thus, each estimate has a standard deviation which is $O(n^{-1/2})$. The order of the standard deviation for an estimate is the reason for neglecting all terms involving n to orders $n^{-3/2}$ and higher in the expected value expressions for these estimates.

4. Numerical Example. To illustrate use of the methods of this paper, let us consider the case where $n = 20$, $x_1, \cdots, x_{15}$ are given, $p = 0.84$, and $R = 17$. The value of m is determined first. This value is the smallest integer which is at least 3 and such that

$$m \geqq \begin{cases} \dfrac{\log [\max (n^{-3/2}, 10^{-4})]}{\log |\frac{1}{2}(p - p_t)|} \doteq 1.84, & \text{if } \theta_p \text{ estimated,} \\ \dfrac{\log [\max (n^{-3/2}, 10^{-4})]}{\log |\frac{1}{2}(p_R - p_t)|} \doteq 1.71, & \text{if } x_R \text{ estimated.} \end{cases}$$

Thus $m = 3$ for both types of estimates.

Next let us evaluate $r(1)$, $r(2)$, and $r(3)$. The value of K is given by

$$K = \max [1, \text{largest integer contained in } \frac{1}{m}\sqrt{n+1}] = 1.$$

Hence, for both types of estimates

$$r(1) = 13, \qquad r(2) = 14, \qquad r(3) = 15,$$

since $r(i) = r - (m - i)K$. Thus $t = 14$ and

$$d(1) = 1, \qquad d(2) = 0, \qquad d(3) = -1.$$

Also the relations

$$\max_{1 \leq i \leq m} |d(i)| \leqq \tfrac{3}{4} p_t q_t (n+1),$$

$$|p - p_t| \leqq \tfrac{1}{4} \min (pq, p_t q_t),$$

$$|p_R - p_t| \leqq \tfrac{1}{4} \min (p_R q_R, p_t q_t)$$

are easily verified so that the methods of the paper are applicable for the case considered.

By direct substitution, the values of the A_j and B_j are found to be

$$A_1 = 1.0000, \quad A_2 = 0.0867, \quad A_3 = 0.0038,$$
$$B_1 = 1.0000, \quad B_2 = 0.0715, \quad B_3 = 0.0028.$$

Thus

$$\begin{aligned} C(1,1) &= 1.0000, & C(1,2) &= 1.0000, & C(1,3) &= 1.0000, \\ C(2,1) &= -0.0390 & C(2,2) &= -0.0867 & C(2,3) &= -0.1344, \\ C(3,1) &= 0.0054 & C(3,2) &= 0.0088 & C(3,3) &= 0.0144, \end{aligned}$$

and

$$\begin{aligned} C'(1,1) &= 1.0000, & C'(1,2) &= 1.000, & C'(1,3) &= 1.0000, \\ C'(2,1) &= -0.0238 & C'(2,2) &= -0.0715 & C'(2,3) &= -0.1192, \\ C'(3,1) &= 0.0049 & C'(3,2) &= 0.0076 & C'(3,3) &= 0.0125. \end{aligned}$$

Consequently,

$$a_1 = \begin{vmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.0867 & -0.0867 & -0.0390 \\ 0.0038 & 0.0088 & 0.0054 \end{vmatrix} \Big/ \begin{vmatrix} 1.0000 & 1.0000 & 1.0000 \\ -0.1344 & -0.0867 & -0.0390 \\ 0.0144 & 0.0088 & 0.0054 \end{vmatrix}$$
$$= 3.38.$$

Similarly, $a_2 = -9.41$ and $a_3 = 7.03$. Also

$$b_1 = \begin{vmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.0715 & -0.0715 & -0.0238 \\ 0.0028 & 0.0076 & 0.0049 \end{vmatrix} \Big/ \begin{vmatrix} 1.0000 & 1.0000 & 1.0000 \\ -0.1192 & -0.0715 & -0.0238 \\ 0.0125 & 0.0076 & 0.0049 \end{vmatrix}$$
$$= 1.49.$$

In a like fashion, $b_2 = -4.99$ and $b_3 = 4.50$.

Using the values determined for the a_i and b_i, approximate expected value estimates are obtained for $\theta_{0.84}$ and x_{17}. These estimates and their properties are given by

$$\begin{aligned} E(3.38\, x_{13} - 9.41\, x_{14} + 7.03\, x_{15}) &= \theta_{0.84} + O(n^{-3/2}) + O[|\tfrac{1}{2}(p - p_i)|^m] \\ &= \theta_{0.84} + O(0.011) + O(0.00065), \\ E(1.49\, x_{13} - 4.99\, x_{14} + 4.50\, x_{15}) &= E(x_{17}) + O(n^{-3/2}) + O[|\tfrac{1}{2}(p_R - p_i)|^m] \\ &= E(x_{17}) + O(0.011) + O(0.00036). \end{aligned}$$

Here the contribution of order $n^{-3/2}$ seems to be much more important that the contribution of order $|\frac{1}{2}(p - p_i)|^m$ or the contribution of order $|\frac{1}{2}(p_R - p_i)|^m$.

To check the expected value accuracy of these estimates, let us consider the special case of a normal distribution with zero mean and unit variance. Using a table of the standardized normal distribution and the results of [3],

$$E(3.38\,x_{13} - 9.41\,x_{14} + 7.03\,x_{15}) = 0.995, \qquad \theta_{0.84} = 0.996,$$

$$E(1.49\,x_{13} - 4.99\,x_{14} + 4.50\,x_{15}) = 0.888, \qquad E(x_{17}) = 0.921$$

Thus the expected values of these two estimates are in rather close agreement with the true values for the case of normality. This expected value agreement is much closer than is required on the basis of the standard deviations of these estimates. For the standardized normal case,

$$\text{s.d. of } (3.38\,x_{13} - 9.41\,x_{14} + 7.03\,x_{15}) \doteq 1.03$$

$$\text{s.d. of } (1.49\,x_{13} - 4.99\,x_{14} + 4.50\,x_{15}) \doteq 0.72.$$

Due to the moderately small value of n, these standard deviation values do not agree very closely with the asymptotic value of

$$\sqrt{p_i q_i}/\sqrt{n} f(\theta_{p_i}) \doteq 0.33.$$

The moderately small value of $n = 20$ was selected for the example in order that the results of [3] could be used.

5. Derivations. Here verification is presented for the expected value and variance results stated in Section 3. This verification is based on the material presented by David and Johnson in [4].

Let s be a number such that $1 \leqq s \leqq n$ while $f(x)$ has derivatives of all orders at all points where it is defined and is non-zero at all points considered. Some additional notation is used

$$F(X) = \int_{-\infty}^{X} f(x)\,dx, \qquad F(X_s) = p_s, \qquad X_s^{(u)} = \frac{d^u X}{dF^u}\bigg|_{X=X_s},$$

$u = 1, 2, \cdots$. Here $X_s^{(0)} = X_s$ while $X_{(n+1)p} = \theta_p$, whether $(n+1)p$ is an integer or not. On the basis of [4],

$$E[x_{r(i)}] = X_{r(i)} + \frac{p_{r(i)}\,q_{r(i)}}{2(n+2)}\,X_{r(i)}^{(2)} + \frac{p_{r(i)}\,q_{r(i)}}{(n+2)^2}\left[\tfrac{1}{3}(q_{r(i)} - p_{r(i)})\,X_{r(i)}^{(3)} + \tfrac{1}{8}p_{r(i)}\,q_{r(i)}\,X_{r(i)}^{(4)}\right]$$

to terms of order n^{-3}, where $p_{r(i)} = r(i)/(n+1)$, and $q_{r(i)} = 1 - p_{r(i)}$. Also

$$E(x_R) = X_R + \frac{p_R\,q_R}{2(n+2)}\,X_R^{(2)} + O(n^{-2}),$$

on the basis of [4].

The first step of the procedure used for developing the estimate of θ_p consists

in expanding the $E[x_{r(i)}]$ and θ_p about a probability value which is halfway between p_t and p. The value p_t is considered because of the relation

$$E[x_{r(i)}] = \theta_{t/(n+1)} + O(1/\sqrt{n}), \qquad (i = 1, \cdots, m).$$

Taylor series expansion about the midway probability value of $\frac{1}{2}(p + p_t)$ yields desirable convergence properties for both the $E[x_{r(i)}]$ and θ_p. In particular, these expansions have about the same rate of convergence. Similarly, the first step in developing the estimate for x_R consists in expanding the $E[x_{r(i)}]$ and $E(x_R)$ about the midway probability value of $\frac{1}{2}(p_R + p_t)$. Here the probability p_R is considered because of the relation $E(x_R) = \theta_{R/(n+1)} + O(n^{-1})$.

Next let us consider the expansion of $E[x_{r(i)}]$ about the general probability value of $\frac{1}{2}(Q + p_t)$. By use of Taylor series,

$$X_{r(i)}^{(u)} = \sum_{v=0}^{\infty} \frac{(-1)^v}{v!} \left[\frac{\frac{1}{2}(n+1)(Q - p_t) + d(i)}{n+1}\right]^u X_{\frac{1}{2}[(n+1)Q+t]}^{(u+v)},$$

$(u = 0, 1, \cdots)$, since

$$\frac{d^v X_s^{(u)}}{ds^v} = \frac{X_s^{(u+v)}}{(n+1)^v}, \qquad (v = 1, 2, \cdots).$$

Substitution of these relations into the expression given for $E[x_{r(i)}]$ shows that

$$E[x_{r(i)}] = \sum_{j=1}^{m} C_j[d(i), \tfrac{1}{2}(Q - p_t)]\, X_{\frac{1}{2}[(n+1)Q+t]}^{(j-1)} + O(n^{-2}) + \sum_{j=m-3}^{m} O\{n^{-(m-j)/2} \,|\, \tfrac{1}{2}(Q - p_t)\,|^j\}.$$

Here the $O(n^{-2})$ and/or the $O[|\,\frac{1}{2}(Q - p_t)|^m]$ terms are the most important of those which are not explicitly stated.

Now the expansions for θ_p and $E(x_R)$ are considered. The Taylor series expansion of θ_p about the probability value of $\frac{1}{2}(p + p_t)$ is

$$\theta_p = \sum_{u=0}^{\infty} \frac{1}{u!} [\tfrac{1}{2}p - p_t)]^u\, X_{\frac{1}{2}[(n+1)p+t]}^{(u)}.$$

To obtain the expansion of $E(x_R)$ to $O(n^{-2})$, the Taylor series expansion of the $X_R^{(u)}$ about the probability value of $\frac{1}{2}(p_R + p_t)$ is needed. This is given by

$$X_R^{(u)} = \sum_{v=0}^{\infty} \frac{1}{v!} [\tfrac{1}{2}(p_R - p_t)]^v\, X_{\frac{1}{2}(R+t)}^{(u+v)}, \qquad (u = 0, 1, \cdots).$$

Substitution of these relations into the expression given for $E(x_R)$ shows that

$$E(x_R) = \sum_{u=0}^{\infty} \left\{\frac{[\frac{1}{2}(p_R - p_t)]^u}{u!} + \frac{p_R q_R u(u-1)}{2(n+2)u!} [\tfrac{1}{2}(p_R - p_t)]^{u-1}\right\} X_{\frac{1}{2}(R+t)}^{(u)}$$

plus terms of order n^{-2}.

To determine the equations which are used to evaluate $a_1, \cdots, a_m$, first

set Q equal to p. Then the coefficient of $X^{(u)}_{[(n+1)p+t]}$ in the expansion for

$$E(\sum a_i x_{r(i)})$$

is required to be the same as the coefficient of this quantity in the expansion of θ_p, $(u = 0, 1, \cdots, m-1)$, to terms of the prescribed order. Examination of the expansions for $E[x_{r(i)}]$ and θ_p shows that the m linear equations in m unknowns given in Section 3 for evaluating the a_i satisfy this requirement when $m \geqq 3$. If $m \geqq 4$, the terms of order $n^{-3/2}$ are also cancelled out.

To determine the equations used to evaluate $b_1, \cdots, b_m$, set Q equal to p_R. Then the coefficient of $X^{(u)}_{[(R+t)]}$ in the expansion for $\mathrm{E}(\sum b_i x_{r(i)})$ is required to be the same as the coefficient of this quantity in the expansion for $E(x_R)$, $(u = 0, 1, \cdots, m-1)$, to terms of the prescribed order. Examination of the expansions for $E[x_{r(i)}]$ and $E(x_R)$ shows that the equations given in Section 3 for evaluating the b_i satisfy this requirement when $m \geqq 3$. If $m \geqq 4$, the terms of order $n^{-3/2}$ also cancel out.

Finally let us consider the variance expressions for the type of estimates considered. Let the $d(i)$ be numbered so that $d(1) > d(2) > \cdots > d(m)$. Then, using the variance results presented in [4] and the general notation c_i to represent either the a_i or the b_i,

$$\begin{aligned}\operatorname{var}\left[\sum_{i=1}^{m} c_i x_{r(i)}\right] &= \sum_{i=1}^{m} \frac{c_i^2}{n}\left[p_i q_i + \frac{(p_i - q_i)\,d(i)}{n+1}\right]\left[X_i^{(1)} - \frac{d(i)}{n+1} X_i^{(2)}\right]^2 \\ &\quad + 2\sum_{i>j=1}^{m} \frac{c_i c_j}{n}\left[p_i q_i + \frac{p_i\,d(i) - q_i\,d(j)}{n+1}\right]\left[X_i^{(1)} - \frac{d(i)}{n+1} X_i^{(2)}\right] \\ &\quad \cdot \left[X_i^{(1)} - \frac{d(j)}{n+1} X_i^{(2)}\right] + O(n^{-2}).\end{aligned}$$

This follows from the fact that all of the a_i, b_i, and $d(i)/\sqrt{n+1}$ are $O(1)$ with respect to n. Using the condition $\sum_1^m c_i = 1$, which holds in all cases for the estimates derived, it is easily verified that

$$\operatorname{var}\left[\sum_{i=1}^{m} c_i x_{r(i)}\right] = \frac{p_i q_i}{n[f(\theta_{p_i})]^2} + O(n^{-3/2});$$

here the relation $X_i^{(1)} = 1/f(\theta_{p_i})$ is used.

REFERENCES

[1] Paul H. Jacobson, "The relative power of three statistics," *Jour. Amer. Stat. Assoc.*, Vol. 42 (1947), pp. 575–84.

[2] B. Epstein and M. Sobel, "Life testing," *Jour. Amer. Stat. Assoc.*, Vol. 48 (1953), pp. 486–502.

[3] D. Teichroew, "Tables of expected values of order statistics and products of order statistics for samples of size twenty or less from the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1955), pp. 410–26.

[4] F. N. David and N. L. Johnson, "Statistical treatment of censored data. Part I—fundamental formulae," *Biometrika*, Vol. 41 (1954), pp. 228–40.

PROBABILITY DISTRIBUTIONS OF RANDOM VARIABLES ASSOCIATED WITH A STRUCTURE OF THE SAMPLE SPACE OF SOCIOMETRIC INVESTIGATIONS[1]

By Leo Katz and James H. Powell

Michigan State University and Western Michigan College

1. Summary. In this paper, we consider a disjoint decomposition, at three levels, of the total sample space for n-person, one-dimensional sociometric investigations. This results in a structure particularly suited to determination of the probability distributions of a large class of sociometric variables. Systematic methods for obtaining these distributions are presented and illustrated by two examples; while the first is trivial, the second produces a previously unknown result.

It should be remarked that the methods developed here have application in the theory of communication networks and, indeed, in the study of any network situations which may be represented by either of the two models employed in the paper.

2. Introduction. The simplest model for the organization of a group of individuals is one-dimensional, in the sense that organization for only one activity of the group is considered. Connections between *ordered* pairs of individuals are represented by non-reflexive binary relations. Although a binary model appears superficially to be too barren to show adequately the richness of variability of the response of one individual to another, it is by no means trivial and is precisely the model used in most sociometric investigations, where the relations are lines of communication, authority, liking, etc.

In this model, a particular organization of n individuals has two isomorphic representations, both of which have been used extensively in the literature for descriptive purposes. The older of the two is the linear directed graph on n points, $P_1, P_2, \cdots, P_n$. A connection from man i to man j is represented by a directed line from P_i to P_j, $P_i \rightarrow P_j$; the absence of such a connection, by no line from P_i to P_j. The equivalent matrix representation is an $n \times n$ matrix, $C = (c_{ij})$, where $c_{ij} = 1$ if a connection exists from man i to man j, and $c_{ij} = 0$, otherwise. By convention, $c_{ii} = 0$. Obviously, $c_{ij} = 1$ (or 0) if and only if a directed line exists (or doesn't) from P_i to P_j. Hence, the two representations are isomorphic.

To fix the notation, let $r_i = \sum_j c_{ij}$ be the ith row total of C and $s_j = \sum_i c_{ij}$ be the jth column total. In the graph, r_i is the number of lines issuing from the point P_i, and s_j is the number of lines terminating on the point P_j. Moreover, $\sum_i r_i = \sum_j s_j = t$, the total number of directed lines. Finally, let the vectors

Received May 24, 1956; revised October 18, 1956.

[1] This work was supported by the Office of Naval Research under contract NR 170-115

r and s, with elements r_i and s_j, respectively, be the two n-part, non-negative, ordered partitions of t which represent respectively, the marginal row and column totals of C.

Unless otherwise noted in the sequel, all graphs will be on n points and linearly directed (n-graphs), and all matrices will be $n \times n$ hollow matrices of 1's and 0's. (A matrix is *hollow* if all principal diagonal elements vanish.)

3. Decomposition of the sample space. The sample space of the possible organizations of an n-member group is the space of all possible n-graphs or $n \times n$ hollow matrices of 1's and 0's. In this section, we consider a decomposition of the total sample space, Ω, following lines which hold promise of utility for certain investigations. We define first-order disjoint subspaces, Ω_t, $t = 0, 1, \cdots, [n(n-1)]$, as the collections of n-graphs containing exactly t lines. Obviously,

$$\Omega = \bigcup_{t=0}^{n(n-1)} \Omega_t, \tag{1}$$

since the Ω_t are mutually exclusive and exhaustive.

Continuing in the same vein, we define second-order subspaces, $\omega(\rho)$, $\rho = (r_1, r_2, \cdots, r_n)$, as the collections of graphs with r_i lines emanating from P_i, $i = 1, 2, \cdots, n$. Since $\sum_i r_i = t$, we have

$$\Omega_t = \bigcup_{(\rho)_t} \omega(\rho), \tag{2a}$$

where $(\rho)_t$ is a generic symbol for non-negative, integral, ordered, n-part partitions of t with all $r_i < n$. In a completely dual manner, we might define alternative second-order subspaces, $\omega(\sigma)$ in terms of n-graphs with s_j lines converging on P_j. In this case, we would have

$$\Omega_t = \bigcup_{(\sigma)_t} \omega(\sigma). \tag{2b}$$

Third-order subspaces are defined by $\omega(\rho, \sigma) \equiv \omega(\rho) \cap \omega(\sigma)$, and are identified with spaces of n-graphs with r_i lines emanating from, *and* s_i lines converging on, P_i. Once again, these sets are exclusive and exhaustive in the sense that

$$\omega(\rho) = \bigcup_{(\sigma)_t} \omega(\rho, \sigma), \tag{3a}$$

and

$$\omega(\sigma) = \bigcup_{(\rho)_t} \omega(\rho, \sigma). \tag{3b}$$

We remark that double and triple disjoint decompositions of the larger spaces may also be indicated.

It will be obvious to the reader that there exist isomorphisms among certain of these second and third-order subspaces. It will be less obvious, but important for computations, that these isomorphisms involve *simultaneous* permutations on the elements of the two vectors ρ and σ. We shall not elaborate on this point since it contributes little to the notions with which we are here concerned.

4. Random variables associated with the structure of the sample space. The decomposition described in the previous section imposes a structure on the sample

space. In most sociometric investigations, involving randomness in the existence of connections between ordered pairs of individuals, it has been deemed appropriate to assign uniform probability to each of the points in a third-order subspace, at least. In more extreme cases (the vast majority) it is customary to assume that every possible sample point is equally likely. Sometimes this has been done without even specifying which sample points are possible under the conditions of the experiment.

In the context of the particular experiment, it is usually possible for the experimenter to determine that his universe of discourse consists of Ω or one of the smaller subsets we have described. If, *also*, it happens that the random variable under discussion assumes the same value over all the points of each of certain smaller subspaces, the assumption of uniformity of probability within these subspaces will produce the complete probability distribution of the variable. In this section, we investigate these circumstances.

We say that a random variable defined over n-graphs or $n \times n$ hollow matrices is *associated* with the sample space structure of the previous section if the value of the variable is constant over all points in every $\omega(\rho, \sigma)$ contained in the domain of definition of the variable. Every such variable has a probability distribution which is completely specified as soon as we are able to count the numbers of points in the appropriate subspaces, assuming uniformity of probability on each point. In the next section, we shall present methods for carrying out this enumeration. A variable associated with the structure in the sense of the present definition is necessarily one whose value is somehow determined by, i.e., is a function of, the r_i and s_j, $i, j = 1, 2, \cdots, n$, alone. Indeed, this may be taken as an alternative definition.

To establish that the class of variables associated with the structure has some real substance, we examine a few variables which have been the subjects of sociometric investigations. Gross expansiveness, or average level of expansiveness, has been defined in terms of t alone in the context of the space Ω. Variability in expansiveness is defined as a function (usually a sum of squares) of the r_i, $i = 1, 2, \cdots, n$, sometimes in the context of Ω and sometimes in Ω_t. A number of variables have been defined as functions of the s_j, $j = 1, 2, \cdots, n$, in various contexts ranging down to $\omega(\rho)$. Examples are (1) the number of isolates, i.e., the number of $s_j = 0$ and (2) the choice status of the most highly chosen, i.e., $\max_j s_j$. Both of these are usually studied in the context of some $\omega(\rho)$.

5. Enumeration of the points in various subspaces. In considering the problems of enumeration, it will be more convenient to use the matrix representation because of its more flexible notation. Thus, the total number of matrices (graphs) in Ω is the number of ways in which the $n(n-1)$ elements of C may be specified as either zero or one. By elementary considerations, the number of distinct ways this can be done is

$$\eta = 2^{n(n-1)}. \tag{4}$$

The matrices in Ω_t have t ones distributed over $n(n-1)$ positions; the number of ways this can be accomplished is the number of ways of specifying a particular t of the $n(n-1)$ positions. Therefore, the number of matrices (graphs) in Ω_t is given by

$$\eta_t = \binom{n(n-1)}{t}. \tag{5}$$

where $\binom{a}{b}$, $b \leqq a$, is the binomial coefficient $a!/[b!(a-b)!]$. As is well-known,

$$\sum_t \eta_t = \sum_t \binom{n(n-1)}{t} = 2^{n(n-1)} = \eta.$$

The enumeration of matrices in $\omega(\rho)$ is accomplished by considering, for each i, r_i ones distributed over $(n-1)$ positions. This can be done, independently, for each i, in $\binom{n-1}{r_i}$ ways and thus the total number of matrices (graphs) in $\omega(\rho)$ is given by

$$\eta(\rho) = \prod_1^n \binom{n-1}{r_i}. \tag{6a}$$

By a similar argument, the number of matrices in $\omega(\sigma)$ is given by

$$\eta(\sigma) = \prod_1^n \binom{n-1}{s_j}. \tag{6b}$$

It is easily seen that

$$\sum_{(\rho)_t} \eta(\rho) = \sum_{(\sigma)_t} \eta(\sigma) = \eta_t .$$

The only difficult counting problem arises when we attempt to compute the number of points in $\omega(\rho, \sigma)$. This problem was solved by the authors [3] who showed that this number is given by

THEOREM.

$$\eta(\rho, \sigma) = A\left\{\left[\prod_{i=1}^n (1 + \delta_i)^{-1}\right](\rho, \sigma)\right\}.$$

where the δ_i are operators on the pair of vectors defined by $\delta_i(r_1, \cdots, r_i, \cdots, r_n ; s_1, \cdots, s_i, \cdots, s_n) = (r_1, \cdots, r_i - 1, \cdots, r_n ; s_1, \cdots, s_i - 1, \cdots, s_n)$, the symbol $A\{\sum a_\alpha(\rho_\alpha, \sigma_\alpha)\}$ stands for $\sum a_\alpha A(\rho_\alpha, \sigma_\alpha)$ and $A(\rho_\alpha, \sigma_\alpha)$ is the coefficient of the monomial symmetric function of order corresponding to σ_α in the expansion of the unitary (elementary) symmetric function of order corresponding to ρ_α.

We note that the coefficients $A(\rho_\alpha, \sigma_\alpha)$ are given in tables of David and Kendall [1] for ρ_α and σ_α partitions of t up to $t = 12$. P. V. Sukhatme [5] gave an algorithm for computing $A(\rho_\alpha, \sigma_\alpha)$ for any weight and showed that $A(\rho, \sigma)$ is the number of matrices of elements $c_{ij} = 0$ or 1 with fixed row totals r_i and

column totals s_j but *without* restrictions on the diagonal elements. We present a very much abbreviated alternative to the proof previously given by the authors in the paper cited above.[2]

PROOF. $G_n \equiv \prod_{i,j=1}^{n} (1 + x_i y_j)$ generates the $A(\rho_\alpha , \sigma_\alpha)$ as coefficients of terms $\prod_{i=1}^{n} x_i^{r_{i\alpha}} \prod_{j=1}^{n} y_j^{s_{j\alpha}}$, and we may write

$$G_n = \sum_{(\sigma)} A(\rho_\alpha , \sigma_\alpha) \prod_i x_i^{r_{i\alpha}} \prod_j y_j^{s_{j\alpha}},$$

where the sum extends over all α such that

$$0 \leqq r_{i\alpha} \leqq n - 1, \qquad 0 \leqq s_{j\alpha} \leqq n - 1, \qquad \sum_i r_{i\alpha} = \sum_j s_{j\alpha} .$$

This is most easily seen if each c_{ij} in a matrix C of 0's and 1's is represented as $(x_i y_j)^{c_{ij}}$. Then, each term in the formal expansion of G_n represents one complete configuration of all the c_{ij}, simultaneously. Finally, in each individual term, the total exponent of $x_i(y_j)$ is the sum $\sum_j c_{ij} = r_i \left(\sum_i c_{ij} = s_j\right)$, and the coefficient $A(\rho_\alpha , \sigma_\alpha)$ is the number of distinct configurations of the c_{ij} with the indicated row and column totals.

Minor modification of the same reasoning serves to establish that

$$H_n \equiv \prod_{\substack{i,j=1 \\ i \neq j}}^{n} (1 + x_i y_j)$$

is a generating function for the $\eta(\rho_\alpha , \sigma_\alpha)$.

Next, we observe that

$$H_n = \left[\prod_{i=1}^{n} (1 + x_i y_i)^{-1}\right] G_n .$$

In this equation, the coefficient of $\left(\prod_i x_i^{r_i} \prod_j y_j^{s_j}\right)$ in the left-hand member is $\eta(\rho, \sigma)$ and, in the right-hand member, is $\sum_{\alpha_1} \sum_{\alpha_2} \cdots \sum_{\alpha_n} (-)^{\sum_i \alpha_i} A(\rho - \alpha, \sigma - \alpha)$, where the α_i range over all non-negative integers. Equating these coefficients gives the expanded form of the statement of the theorem.

We note that the last sum in the proof above may be written in finite terms, since, as soon as any $\alpha_i > \min (r_i , s_i)$, the corresponding $A(\rho - \alpha, \sigma - \alpha) \equiv 0$, by the definition of Sukhatme as a number of certain matrices of 0's and 1's.

6. Probability distributions of associated random variables. It is now clear that we have laid down a program for computing, exactly, the probability distributions for any and all random variables associated with this structure of the sample space. In particular instances, it may be possible to effect certain economies in the computations by exploiting the isomorphisms among subsets so as to avoid duplication.

When the variable in question has constant values on sub-spaces no larger than an $\omega(\rho, \sigma)$, the computations are always formidable, though never impossible. In such circumstances, it would seem desirable to develop approximate

[2] This alternative proof follows lines of a suggestion by J. S. Frame

distributions for these variables, treating the exact methods as procedures for testing the validity of the approximations over the ranges of group size, etc., to be covered. For *very* small groups, it will usually be feasible to carry out the exact computations.

7. Examples. We shall give two examples of random variables associated with the sample space structure. In each, we consider the null case in which each graph in the appropriate sample space is equally likely, i.e., a uniform probability distribution over the sample space.

EXAMPLE 1. One measure of gross expansiveness, equal to the total number of choices made by group divided by size of group, is given by Loomis and Proctor [4] in a contribution to *Research Methods in Social Relations*. In our notation, this index is $E = t/n$.

The distribution problem, in the null case, is easily solved. Clearly, the appropriate sample space is Ω and our random variable, the number of distinct n-graphs with t lines, is constant over the first-order subspaces, Ω_t, in the disjoint and exhaustive decomposition of Ω. Thus, our random variable is associated with the sample space structure and according to Section 4 and the enumeration formulas of Section 5, the required probabilities are given by

$$P(t = k) = \frac{\eta_k}{\eta} = \frac{\binom{n(n-1)}{k}}{2^{n(n-1)}}. \tag{7}$$

EXAMPLE 2. An *isolate* is an individual represented in the graph by a point, P_i, with no terminating lines and in the matrix by a column of zeros, i.e., $s_i \parallel 0$ in the vector σ. The exact probability distribution of the number of isolates for the case $r_i = d(i = 1, 2, \cdots, n)$ was obtained from first principles by Katz [2], in 1950.

Using the methods already developed, we can now easily extend this result to the general case where the ith individual has r_i outgoing connections, the r_i being not necessarily equal.

The most common setting for this problem is in the sample space $\omega(\rho)$. In the null case, we desire the number of n-graphs having a specified number of points with no terminating lines, i.e., a specified number of zeros in the vector σ. Our random variable, X, the number of zero s_j's, is constant over the third order subspaces $\omega(\rho, \sigma)$ in the decomposition of $\omega(\rho)$; thus, it is associated with the sample space structure. Hence, according to Section 4 and the enumeration formulas of Section 5, the probability of exactly k isolates is given by

$$\begin{aligned} P(X = k \mid \rho) &= \frac{\sum_{(\sigma)_k} I_{A_k}(\sigma)\eta(\rho, \sigma)}{\eta(r)} \\ &= \frac{\sum_{(\sigma)_k} I_{A_k}(\sigma) A\left\{\prod_1^n (1 + \delta_i)^{-1}(\rho, \sigma)\right\}}{\prod_1^n \binom{n-1}{r_i}}, \end{aligned} \tag{8}$$

where A_k is the union of $\omega(\rho, \sigma)$ such that the vectors σ have exactly k vanishing components, and I_A is the indicator function for the set A.

We remark that in some contexts the appropriate sample space might be the larger space Ω_t. However, our enumeration methods will still give us the required probabilities necessary to construct the distribution. In this case, the probability of exactly k isolates is given by

$$P(X = k \mid t) = \frac{\sum_{(\rho)_t} \sum_{(\sigma)_t} I_{A_k}(\sigma)\eta(\rho, \sigma)}{\eta_t}$$

$$= \frac{\sum_{(\rho)_t} \sum_{(\sigma)_t} I_{A_k}(\sigma) A\left\{\prod_1^n (1 + \delta_i)^{-1}(\rho, \sigma)\right\}}{\binom{n(n-1)}{t}}, \tag{9}$$

where the notations are the same as before.

Thus, the probability distribution can be constructed for any index (proposed for the study of group structure) which depends only on the number of isolates in the group. Another such index, equal to the reciprocal of the number of isolates, is given by Loomis and Proctor [4] as a measure of "group integration."

Finally, we note that neither of the distributions (8) and (9) have been given correctly in the literature.

REFERENCES

[1] F. David and M. G. Kendall, "Tables of symmetric functions. Parts II and III," *Biometrika*, Vol. 38 (1951), pp. 435–462.

[2] L. Katz, "The distribution of the number of isolates in a social group," *Ann. Math. Stat.*, Vol. 23 (1951), pp. 271–276.

[3] L. Katz and J. Powell, "The number of locally restricted directed graphs," *Proc. Amer. Math. Soc.*, Vol. 5 (1954), pp. 621–626.

[4] C. Loomis and C. Proctor, "Analysis of Sociometric Data," *Research Methods in Social Relations*, Part 2, Chap. 17, The Dryden Press, New York, 1951.

[5] P. V. Sukhatme, "On bipartitional functions," *Philos. Trans. Roy. Soc. London, Ser. A*, Vol. 237 (1938), pp. 375–409.

DESIGN FOR THE CONTROL OF SELECTION BIAS[1]

By David Blackwell and J. L. Hodges, Jr.

University of California, Berkeley

0. Summary. Suppose an experimenter E wishes to compare the effectiveness of two treatments, A and B, on a somewhat vaguely defined population. As individuals arrive, E decides whether they are in the population, and if he decides that they are, he administers A or B and notes the result, until nA's and nB's have been administered. Plainly, if E is aware, before deciding whether an individual is in the population, which treatment is to be administered next, he may, not necessarily deliberately, introduce a bias into the experiment. This bias we call *selection bias*. We propose to investigate the extent to which a statistician S, by determining the order in which treatments are administered, and not revealing to E which treatment comes next until after the individual who is to receive it has been selected, can control this selection bias.

Thus a design d is a distribution over the set T of the $\binom{2n}{n}$ sequences of length $2n$ containing nA's and nB's. We shall measure the bias of a design by the maximum expected number of correct guesses which an experimenter can achieve, knowing d, attempting to guess the successive elements of a sequence $t \varepsilon T$ selected by d, and being told after each guess whether or not it is correct. The distribution of the number G of correct guesses depends both on d and on the prediction method p used by the experimenter. We shall consider particularly two designs, the *truncated binomial*, in which the successive treatments are selected independently with probability $\frac{1}{2}$ each until n treatments of one kind have occurred, and the *sampling* design, in which all $\binom{2n}{n}$ sequences are equally likely. We shall consider particularly two prediction methods, the *convergent* prediction, which predicts that treatment which has hitherto occurred less often, and the *divergent* prediction, which predicts that treatment which has hitherto occurred more often, except that after n treatments of one kind have been administered, the divergent prediction agrees with the convergent predictions that the other treatment will follow; when both treatments have occurred equally often, either method predicts A or B by tossing a fair coin, independently for each case of equality.

We find that among all designs, the truncated binomial minimizes the maximum expected number of correct guesses. For this design, the expected number of correct guesses is independent of the prediction method, and is

$$n + n\binom{2n}{n} \Big/ 2^{2n} \sim n + (n/\pi)^{1/2}.$$

Received March 26, 1956; revised November 2, 1956.

[1] This investigation was supported (in part) by research grant from The National Institutes of Health, Public Health Service.

With the truncated binomial design, the variance in the number of correct guesses is largest for the divergence strategy and is

$$3n/2 - D - D^2/4 \sim (3\pi - 2)n/2\pi - 2(n/\pi)^{1/2},$$

where $D = n\binom{2n}{n} \Big/ 2^{2n-1}$, and is smallest for the convergence strategy, and is $n/2 - D^2/4 \sim (\pi - 1)n/2\pi$. For the sampling design, convergent prediction maximizes the expected number of correct guesses; this maximum is

$$n + 2^{2n-1} \Big/ \binom{2n}{n} - \frac{1}{2} \sim n + (\pi n/4)^{1/2}.$$

Finally we note that, if treatments are selected independently at random, bias of the kind we discuss disappears, but the treatment numbers can no longer be preassigned. Three such designs are considered: the *fixed total* design, in which the total number of treatments is a fixed number s, the *fixed factor* design, in which we continue until $1/X + 1/Y \leqq 2/n$, where X is the number of A treatments and Y is the number of B treatments administered, and the *fixed minimum* design, in which we continue until $\min(X, Y) = n$. For the fixed total design, we find that, for $s = 2n + 4$, $\Pr(1/X + 1/Y \leqq 2/n) \sim 0.955$ for large n; at the expense of 4 extra observations, we have a bias-free design whose variance factor will with probability 0.955 be smaller than that in which treatment numbers are preassigned. For the fixed factor design, the additional number of observations required to achieve the given precision has for large n the distribution of the square of a normal deviate. For the fixed minimum design, in which we guarantee precision for the estimated effect of each treatment, the expected number of additional observations is roughly $1.13\ (n)^{1/2}$.

1. Introduction. It is widely recognized that experiments intended to compare two or more treatments may yield biased results if the experimental subjects are selected with knowledge of the treatments they are to receive. Consider as illustration of experiment in cloud seeding. From a sequence of storms the meteorologist selects $2n$ storms deemed suitable for seeding. Of these, n are seeded and we compare the rainfalls they produce with those produced by the other n storms. If the meteorologist knows (or can guess), while considering the suitability of a storm, whether or not the storm will be seeded if he selects it, there exists the possibility that his selection will be biased.

We shall call this *selection bias*. It presents a serious problem when the trials constituting the experiment occur sequentially in time. If it were possible to collect at one time a block of as many subjects as there are treatments, a simple random assignment of treatments to subjects would dispose of the bias. But in many experiments potential subjects occur singly and must be dealt with when they arise. For example, in clinical trials it is often essential to treat the patient as soon as the illness is diagnosed—the physician cannot wait until he has a similar patient merely to permit randomization of the bias.

In some cases it is possible to eliminate selection bias by conducting the ex-

periment in such a way that the person who selects the subjects is not otherwise involved in the experiment or is not able to discover which treatments have been applied. Again, it may be possible to define subject suitability with precision and to accept without conscious selection all subjects meeting the criteria. But oftentimes the exercise of judgment is essential if the treatments are to have a convincing test, and the best or only judges available are those most deeply involved in administering the treatments. Therefore we thought it interesting to see to what extent selection bias can be controlled through design of the experiment—i.e., through the statistician's strategy in choosing the sequence of treatments to be given to the subjects selected by the experimenter.

Admittedly, selection bias will usually operate subconsciously, but to sharpen the problem we imagine an experimenter E who is consciously seeking to produce biased experimental results, while the statistician S is attempting to prevent this. To fix the problem, suppose we wish to compare two treatments, say A and B. It is customary to decide in advance of the experiment how many subjects will receive each treatment, and it is also customary to assign equal numbers of subjects to the two treatments. While we shall return to this question below, at first let us suppose it given that each treatment will be administered to n subjects.

If E wishes to make it appear that A produces a greater response X than does B, and if he knows (or guesses) that S will assign treatment A to the next subject, then E will try to select a subject whose expected response $E(X)$ is high. Conversely, if E anticipates a B treatment, he will select a subject with low $E(X)$. The results of E's guesses and S's assignments can be displayed in a two-by-two table:

		Number of times when S assigns	
		A	B
and E guesses	A	α	$n - \beta$
	B	$n - \alpha$	β

Suppose that the treatment effects do not differ, but that when E anticipates an $A(B)$ treatment he selects a subject with expected response $\mu + \Delta(\mu - \Delta)$. Then the expected difference of treatment sums is

$$2\Delta(\alpha + \beta - n) = 2\Delta(G - n). \tag{1.1}$$

The quantity G is thus the total number of correct guesses. If he guesses at random, E would on the average be right half the time, giving $E(G - n) = 0$. His ability to bias the experiment depends on getting $E(G)$ above n.

In accordance with the foregoing analysis we now formulate our design problem as a two-person game. The game is played in $2n$ moves. On each move, each of the players E and S privately selects one of the letters A and B, with the restriction that exactly n of S's choices must be A. They then compare

selections; if they agree S pays E one unit. The total payoff is G, and we wish to know S's minimax strategy for minimizing $E(G)$, i.e. the optimum design for controlling selection bias. The value of the game will indicate to what extent selection bias can be controlled through design.

2. The biases of three designs. Before giving the solution of our game, we shall illustrate the ideas by deriving the optimum strategies for E and the corresponding biases for three designs used in experimental work.

(i) A very common practice is the alternation of treatments, producing the treatment sequence $ABAB \cdots AB$ or $BABA \cdots BA$. While this design is exceedingly simple and does an optimum job of spreading the treatments over time, it is about the worst possible design from the standpoint of selection bias. Since E can correctly guess every treatment, $E(G) = 2n$. Even if, as is sometimes done, S selects one of the two patterns at random, E can guess all but the first trial and has half a chance for that, so that $E(G) = 2n - \frac{1}{2}$. (Exactly the same conclusions apply to the "Student" sandwich design $ABBAABBA \cdots ABBA$).

(ii) As just remarked, S can insure that E's expectation of correct guessing on the initial trial is only $\frac{1}{2}$ by simply choosing a treatment at random. Further, S cannot do any better than this, since E can guarantee himself half a chance, whatever S may do, by guessing at random. A similar analysis applies to the second trial, and to all trials until one treatment has been given to n subjects. At that point the requirement that each treatment be given to just n subjects takes over, and the remainder of the subjects must be given the unexhausted treatment. We shall refer to this as the *truncated binomial* design.

Suppose S has announced that he will adopt the truncated binomial design. What should E do, and how large can he make $E(G)$? In the tail of the experiment, consisting of the terminal sequence of trials having like treatment, E knows which treatment will be assigned, so he is sure to guess all of these correctly—let R denote the number of trials in the tail. We take advantage of the fact that $E(G)$ is independent of E's strategy except in the tail, and give to E a strategy which simplifies the calculation. Suppose E guesses A every time, except of course in a tail of B's. Then G must be at least n (since n A's are used), and may in addition contain a B tail. By symmetry, $E(G) = n + E(R)/2$. We must now discuss the distribution of R.

To calculate the probability that $R = r$, notice that this may occur in two ways: the nth A treatment, or the nth B treatment, is assigned on the $(2n - r)$th trial. These events have equal probability $\binom{2n-r-1}{n-1} \Big/ 2^{2n-r}$ according to the negative binomial distribution. Therefore R has a truncated negative binomial distribution,

$$\Pr(R = r) = \binom{2n-r-1}{n-1} \Big/ 2^{2n-r-1}, \qquad r = 1, 2, \cdots, n. \tag{2.1}$$

By calculating $E(2n - R)$, it is easy to establish

$$E(R) = n\binom{2n}{n}\Big/ 2^{2n-1} \approx \frac{2(n)^{1/2}}{\pi^{1/2}} - \frac{1}{4(\pi n)^{1/2}} + \cdots. \tag{2.2}$$

In a similar way, calculation of $E[(2n - R)(2n - R + 1)]$ yields $E(R^2) = 2n - E(R)$. Furthermore, $R(2n)^{1/2}$ has asymptotically the distribution of the absolute value of a normal deviate. For example, there is about one chance in ten that R will exceed $2.32\ n^{1/2}$. Combining with the result of the previous paragraph, we see that the value of the truncated binomial design is $E(G) = n + n\binom{2n}{n}\Big/ 2^{2n}$. The excess $E(G) - n \approx n^{1/2}/\pi^{1/2} - 1/8(\pi n)^{1/2} + \cdots$ is shown in Table II for a number of values of n.

(iii) In the *random allocation* design, S selects n of the first $2n$ positive integers at random without replacement, and then assigns treatment A to those subjects whose ordinal numbers have been selected. Another way of expressing this strategy is to say that on each move S selects a treatment with probability proportional to the number of subjects still to receive that treatment.

It is intuitively clear that against this strategy, E should always use the *convergence* strategy, i.e., he should guess that treatment which has previously been less used; when there is a tie in past use, S will choose A or B with equal probability so E's choice is arbitrary. In calculating $E(G)$ for the sampling design it is very convenient to picture the results of S's choices as a walk on the lattice points of the plane. We start at the point (0, 0), and move one unit to the right (or up) when S picks treatment A (or B). The experiment terminates when we reach the point (n, n). In terms of this walk, E will always guess that the walk will move toward the diagonal—if the walk is on the diagonal his guess is arbitrary. Since the walk starts and stops on the diagonal, it must move towards it exactly n times and away exactly n times. Therefore E's strategy assures him of n correct guesses. In addition, there will be a number of steps originating on the diagonal, say T of them.

If we denote generically by $B(k)$ the number of successes in k binomial trials of success probability one-half, we see that $G = n + B(t)$ when $T = t$. Thus $E(G \mid T = t) = n + t/2$, and $E(G) = n + E(T)/2$.

The distribution of T has been studied by Feller, and it is apparent from formula (6.15) of Chapter 12 of [1] that

$$\Pr(T = t) = 2^t\left[\binom{2n - t - 2}{n - t} - \binom{2n - t - 2}{n - t - 2}\right]\Big/\binom{2n}{n} \tag{2.3}$$

from which it follows that, for large n, $T/n^{1/2}$ has asymptotically the distribution of an absolute normal deviate. If we consider the probability of the walk passing through the point (j, j), we see that

$$E(T) = \sum_{j=0}^{n-1}\binom{2j}{j}\binom{2n - 2j}{u - j}\Big/\binom{2n}{n}. \tag{2.4}$$

The identity (see [2], p. 252)

$$\sum_{j=0}^{n} \binom{2j}{j}\binom{2n-2j}{n-j} = 2^{2n} \tag{2.5}$$

gives

$$E(T) = 2^{2n} \Big/ \binom{2n}{n} - 1 \approx n^{1/2} - 1 + \frac{\pi^{1/2}}{8(n)^{1/2}} + \cdots. \tag{2.6}$$

The asymptotic approximation has error of 0.03 per cent at $n = 5$.

Table 1 gives some values of $E(G) - n$ computed with the aid of these formulas. Notice that the truncated binomial design has in each case a smaller value of $E(G) - n$ than does the random allocation design. This is not an accident, as we now proceed to show.

3. Solution of the game.

THEOREM 1. *The truncated binomial design is the solution of our game.*

In proving this theorem, it is helpful to generalize the problem to permit different preassigned numbers of subjects for the treatments. Let $D(m, k)$ denote the design problem when we are required to use A just m times and B just k times. By analogy, we say S uses the truncated binomial design if he chooses treatments independently and at random until one of the treatments is exhausted. As in the special case $D(n, n)$ it is easy to see that $E(G)$ does not depend on E's strategy (provided always that he guesses the obvious in the tail) when S uses the truncated binomial strategy. If we denote this invariant value of $E(G)$ by $\phi(m, k)$ for the problem $D(m, k)$, we easily find that

$$\phi(m, 0) = m; \qquad \phi(0, k) = k;$$

$$\phi(m, k) = [1 + \phi(m, k-1) + \phi(m-1, k)]/2 \quad \text{for } m, k > 0.$$

For future reference we note that

$$|\phi(m-1, k) - \phi(m, k-1)| < 1. \tag{3.1}$$

TABLE 1

$E(G) - n$

n	Truncated Binomial Design	Sampling Design
0	1.23	1.53
10	1.76	2.24
15	2.17	2.96
20	2.51	3.49
25	2.81	3.95
30	3.08	4.37
40	3.56	5.12
50	3.98	5.78
100	5.63	8.37
∞	$0.564(n)^{1/2}$	$0.886(n)^{1/2}$

This is obvious for $m + k = 2$, and since ϕ on the line $m + k = s + 1$ is just one-half more than the average of consecutive values on the line $m + k = s$, (3.1) holds in general.

Our design problems D are inductively related. Suppose we have checked our theorem for the design problems $D(m - 1, k)$ and $D(m, k - 1)$, showing that the truncated binomial strategy solves these, yielding values $\phi(m - 1, k)$ and $\phi(m, k - 1)$ respectively. We now consider the game $D(m, k)$. After the first move we shall be faced with one of the former games. Therefore the payoff matrix can be expressed in terms of the choices of E and S on the first move only. In fact, the expected payoffs are given by

		S	
		A	B
E	A	$1 + \phi(m - 1, k)$	$\phi(m, k - 1)$
	B	$\phi(m - 1, k)$	$1 + \phi(m, k - 1)$

Now we hope to show that S should choose the columns with equal probabilities. Therefore, let us try to find a strategy for E which will make these columns equally attractive to S. This leads at once to having E choose the first row with probability

$$[1 + \phi(m, k - 1) - \phi(m - 1, k)]/2. \tag{3.2}$$

(That (3.2) is indeed a probability follows from (3.1)). The game is now solved, since (a) when E uses (3.2), the options are equally attractive to S who is then content to choose them at random, while (b) S's random choice makes E indifferent and hence content with (3.2).

Incidentally, our game has an interesting feature. When either player uses his minimax strategy, the expected outcome of the game is independent of the strategy adopted by the other player. Notice also that we have shown the truncated binomial design to be the solution of the general design problem $D(m, k)$, with preassigned but possibly different treatment numbers.

We remark that although this design minimizes $E(G)$, the minimized value is disturbingly large. If we divide the difference of treatment sums (1.1) by $n^{1/2}$ as is customary in standardizing it, the expected value is about $2\Delta/\pi^{1/2}$, which does not tend to 0 as $n \to \infty$. In many experimental situations Δ could be large enough to produce a serious distortion.

4. The variance of G. When S uses the truncated binomial design, the value of $E(G)$ is independent of E's strategy, but it should not be thought that E is unable to influence other aspects of the distribution of G. For example, if E guesses the treatment A, as long as that treatment is possible, he is assured of at least n correct guesses, while if he guesses at random, G can be as low as 1. We shall in particular examine the influence of E's strategy on the variance $V(G)$. This would be an essential quantity in computing the expectation of a payoff function which can be represented by a quadratic function of G, or in approximating the probability that the estimated treatment effect exceeds a specified

critical value. (If E believes that the treatments are not different, but wishes as large as possible a probability of having the difference appear highly significant, he would want $V(G)$ large.)

Our methods permit us to find the strategy for E which will maximize (minimize) $V(G)$. We have introduced above the convergence strategy, according to which E always guesses the hitherto least frequently used treatment. Opposite to this is the *divergence* strategy: as long as both treatments are available, E guesses the one which has been most used; when there is a tie, he guesses at random; while as always in the tail he guesses the treatment certain to be used.

THEOREM 2. *Against the truncated binomial design, $V(G)$ is maximized (minimized when E uses the divergence (convergence) strategy.*

Since $E(G)$ is constant it will suffice to prove the corresponding assertion for $E(G^2) = V(G) + E^2(G)$. Consider the problem $D(m, k)$ where to avoid obvious cases we assume $mk > 0$. Let E employ the pure strategy of guessing A on the first trial. If S assigns A, E wins 1 on that trial and is faced with the game $D(m - 1, k)$, in which E wins, say, H. If S assigns B, E wins nothing on the first trial and then must play $D(m, k - 1)$, winning K. As the assignments are equally likely,

$$E(G^2) = [E(1 + H)^2 + E(K^2)]/2 = \tfrac{1}{2} + \phi(m - 1, k) + E(H^2 + K^2).$$

Similarly, if E adopts pure strategy B on the first trial,

$$E(G^2) = \tfrac{1}{2} + \phi(m, k - 1) + E(H^2 + K^2).$$

Now the distributions of H and K depend on the strategies adopted in playing $D(m - 1, k)$ and $D(m, k - 1)$ respectively, but not on the guess which E makes in the first trial. Therefore, $E(G^2)$ will be maximized when E guesses A on the first trial if $\phi(m - 1, k) > \phi(m, k - 1)$; and when E guesses B if the inequality is reversed. As $\phi(m, s - m)$, viewed as a function of m, is an increasing function of $|\, m - s/2 \,|$, we see that the divergence strategy will maximize $E(G^2)$. A similar argument shows that $E(G^2)$ is minimized by the convergence strategy.

As argued in Section 2, when E adopts the convergence strategy, and the walk has t ties, $G = n + B(t)$. Thus G has as its distribution a mixture of binomials, the mixing coefficients being given by the distribution of T. We shall derive this by considering first the joint distribution of T and R. Denote

$$\Pr(T = t \quad \text{and} \quad R = r)$$

by $\pi(t, r)$, and observe that these variables have the range $3 \leqq t + r \leqq n + 1$, $1 \leqq t, r$.

It is remarkable that $\pi(t, r)$ depends only on $t + r$. This can be seen by establishing a two-to-one mapping of the walks with values $(t + 1, r)$ onto walks with values $(t, r + 1)$. Consider any walk W with $T = t + 1$, $R = r$. Let W' be the walk identical with W except that the part after the last tie has been reflected about the diagonal; W' also has $T = t + 1$, $R = r$. Each of these walks

has probability $1/2^{2n-r}$ under the truncated binomial design. Now locate on W(or W') the point immediately preceding the last tie, and denote its coordinates by (x, y). We shall assume $y = x + 1$ with the case $x = y + 1$ being argued similarly. Suppose W is the walk having its last part above the diagonal. We now create from W a new walk W^*, by (a) eliminating the step from (x, y) to $(x + 1, y)$, (b) shifting one step to the left that part of W from $(x + 1, y)$ to the boundary, and (c) closing the gap thus created by adding a step to the tail. Note that the correspondence between the pair (W, W') and W^* is one-to-one, that W^* has probability $1/2^{2n-r-1}$, and that it has $T = t$, $R = r + 1$.

As a corollary we observe that T and R are identically distributed. Since we have already obtained the distribution of R (2.1), we can now give that of G:

$$\Pr(G = g) = \sum_{i=1}^{n} \binom{t}{g-n}\binom{2n-t-1}{n-1} \Big/ 2^{2n-1}.$$

$V(G)$ can also easily be calculated. We have

$$E(G^2 \mid T = t) = n^2 + nt + (t + t^2)/4$$

so that $E(G^2) = n^2 + (n + \frac{1}{4})ET + \frac{1}{4}ET^2 = n^2 + n/2 + nE(R)$. Since $E(G) = n + E(R)/2$, we get

$$V(G) = n/2 - E^2(R)/4 \approx \frac{\pi - 1}{2\pi} n + \frac{1}{4\pi} + \cdots. \tag{4.1}$$

This is the smallest value which $V(G)$ can have.

Since $\pi(t, r)$ depends only on the sum of its arguments,

$$\pi(t, r) = \pi(t + r - 1, 1).$$

A walk which has $T = t + r - 1$ and $R = 1$ must have just $t + r - 2$ ties before reaching the point $(n - 1, n - 1)$. Each such walk has probability $\frac{1}{2}^{2n-1}$ and the number of them can be read at once from (2.3). It follows that, for $n > 1$,

$$\pi(t, r) = \left[\binom{2n-t-r-2}{n-t-r+1} - \binom{2n-t-r-2}{n-t-r-1}\right] \Big/ 2^{2n-t-r}. \tag{4.2}$$

We shall need $E(RT)$. If we let U_k indicate a tie at (k, k), so that $T = U_0 + U_1 + \cdots + U_{n-1}$, we see that

$$E(RT) = \sum_{k=0}^{n-1} P(U_k = 1)\, E(R \mid U_k = 1) = \sum_{k=0}^{n-1} (n - k)\binom{2k}{k}\binom{2n-2k}{n-k} \Big/ 2^{2n-1}.$$

Again making use of (2.4), we find that $E(RT) = n$.

In computing $V(G)$ for the divergence strategy, note that when $T = t$ and $R = r$, $G = n - t + r + B(t)$. Therefore,

$$E(G^2 \mid T = t, R = r) = (n + r - t)(n + r) + (t + t^2)/4.$$

Using the relations $E(T) = E(R)$, $E(T^2) = E(R^2) = 2n - E(R)$, and $E(RT) = n$, we find after simplifying

$$(4.3)\quad V(G) = 3n/2 - E(R) - E^2(R)/4 \approx \frac{3\pi - 2}{2\pi}\, n - \frac{2(n)^{1/2}}{\pi^{1/2}} + \frac{1}{4\pi} + \cdots .$$

This is the greatest value which $V(G)$ can attain.

Note that the range of $V(G)$ is quite large. The ratio of maximum to minimum values tends with large n to $(3\pi - 2)/(\pi - 1) = 3.467$.

5. Completely binomial designs. Since it is not possible to find a design with adequate bias control when the treatment numbers are preassigned, we shall now examine some designs free of this restriction. In the present section we shall assume that each subject has probability $\frac{1}{2}$ of receiving each treatment, and that the assignments are independent. Such *completely binomial* designs are bias-free, in the sense that every guessing strategy will produce a G whose expectation, given the number s of trials, is exactly $s/2$. We can still exercise a measure of control over the experiment through the decision to terminate it. In our geometrical picture, the design of the experiment now consists in specifying a set of points in the plane at which experimentation will stop. Each such sequential stopping rule will provide a distribution of the numbers X and Y of subjects receiving treatments A and B, respectively, leading to distributions of the total number of trials $X + Y = S$ and of the variance factor $1/X + 1/Y = V$.

(i) *Fixed total design.* In some experiments it may be necessary or desirable to know in advance the total number s of trials to be performed. This leads to the stopping rule $x + y = s$, for which the variance factor V is variable and indeed unbounded: if x or y is 0, $V = \infty$. However, since $X = B(s)$, if s is large it is unlikely that X will be far from $s/2$ and V will probably not much exceed its minimum value $4/s$. In fact, if we expand V in powers of $(x - s/2)$, we find

$$v = \frac{4}{s} + \frac{16}{s^3}\left(X - \frac{s}{2}\right)^2 + \cdots .$$

Here $2(X - s/2)/s^{1/2}$ is approximately distributed as a normal deviate. If K_α denotes the upper $\alpha/2$ point on the normal distribution, and if we want to have $V \leqq 2/n$ with probability $1 - \alpha$, we should choose s so that

$$\frac{2}{n} = \frac{4}{s} + \frac{4}{s^2} K_\alpha^2 ,$$

or $s = 2n + K_\alpha^2$. For example, if we set $s = 2n + 4$, we shall be for large n about 95.5 per cent sure of obtaining a bias-free experiment with variance factor smaller than that obtainable with $X = Y = n$ preassigned.

(ii) *Fixed factor design.* Instead of fixing S and permitting V to vary, we

might often prefer to fix V and permit S to vary. For example, we could continue taking observations until X and Y satisfy

$$\frac{1}{X} + \frac{1}{Y} \leqq \frac{2}{n}. \tag{5.1}$$

Write $X + Y = S = 2n + U$, so that $U \geqq 0$ may be viewed as the number of additional observations required to obtain freedom from bias.

For a given value $U = u$, let x_u denote the smallest value of X for which (5.1) holds; i.e.,

$$\frac{1}{x_u} + \frac{1}{y_u} \leqq \frac{2}{n} < \frac{1}{x_u - 1} + \frac{1}{y_u + 1}, \qquad x_u + y_u = 2n + u. \tag{5.2}$$

A path will yield $U \leqq u$ if and only if, at the $(2n + u)$th step it is at, say, (X^*, Y^*) with $x_u \leqq X^* \leqq y_u$. As $X^* = B(2n + u)$, distribution of U is now easily calculated.

As $n \to \infty$, U has a simple limit law. From (5.2) it appears that for large n, $y_u \sim (n + u/2) + u^{1/2}(2n + u)^{1/2}/2$. Since the binomial X^* has $EX^* = n + u/2$ and $\sigma_{X^*} = (2n + u)^{1/2}/2$, we have

$$\Pr(U \leqq u) \approx \Phi(u^{1/2}) - \Phi(-u^{1/2}), \qquad u = 0, 1, 2, \cdots.$$

Table 2 compares the distributions of U for $n = 5$, 10, 20, and ∞. We see that on the average it costs about one and one-half observations, and is practically certain not to cost as many as ten observations, to eliminate selection bias entirely. (Even this comparison is unfair to the bias-free design, as the inequality (5.1) is usually strict and we are obtaining a somewhat more accurate estimate. If we were to take the final step with a probability adjusted to make $E(1/X + 1/Y) = 2/n$, we should find $EU \to 1$ as $n \to \infty$.)

TABLE 2

Fixed factor design

Distribution of U = excess observations required

	$P(U \leqq u)$			
u	$n = 5$	10	20	∞
0	.246	.176	.125	0
1	.773	.617	.651	.6827
2	.854	.866	.836	.8427
3	.908	.907	.934	.9167
4	.943	.936	.951	.9545
5	.965	.985	.984	.9747
6	.979	.991	.989	.9857
7	.987	.994	.992	.9918
8	.999	.996	.994	.9953
9	.999	.998	.999	.9973
10	1.000	.999	.999	.9984
$E(U)$	1.338	1.464	1.535	1.6625

In practice, the fixed factor design would be used in a truncated form. For example, we could continue the binomial choice until (5.1) holds, or until $X + Y \geqq 2n + u$. If the latter eventuates first, the deficient treatment could be applied until (5.1) holds. By setting $u = 10$, we would have practical certainty of a bias-free design, without the theoretical possibility of an infinite sequence of trials.

(iii) *Fixed minimum design.* In some cases we might wish to guarantee the precision of estimation for each treatment effect separately, rather than for their difference. We should then need

$$\min (X, Y) = n.$$

By symmetry, we are equally likely to stop at (n, x) and at (x, n), so it will suffice to consider the probabilities of stopping at points (x, n) for $x = n, n + 1, \cdots$. These probabilities are easily seen to be proportional to those of the single negative binomial design, which is stopped by $y = n$. Thus our X has a truncated negative binomial distribution, with range just the complement of that of $n - R$ considered in Section 2. As each of the ranges is equally likely, we must have $\frac{1}{2}ES + \frac{1}{2}E(2n - R)$ equal to the expected number of steps in the single negative binomial, which is $2n$. Therefore $ES = 2n + ER$, where $E(R)$ is given by (2.2). Roughly, we must expend on the average $1.13\ n^{1/2}$ additional observations in this case.

6. Extensions. A good deal of the preceding argument generalizes rather easily to experiments involving more than two treatments. In particular, the minimax design for preassigned treatment numbers consists in choosing at each step among the remaining treatments with equal probabilities. A simple bias-free design, which generalizes 5(ii), consists in choosing a treatment at each step, with equal probabilities, and terminating the experiment when the sum of reciprocals of treatment numbers falls below a preassigned level.

REFERENCE

[1] FELLER, WILLIAM, *An Introduction to Probability Theory and Its Applications*, Vol. 1, John Wiley & Sons, 1950.

ON INFINITELY DIVISIBLE RANDOM VECTORS[1]

By Meyer Dwass and Henry Teicher

Northwestern and Stanford Universities; Purdue and Stanford Universities

1. Summary. A normally distributed random vector X is well known to be representable by $A \cdot Y$ (in the sense of having identical distributions), where A is a matrix of constants and Y is a random vector whose component random variables are independent. A necessary and sufficient condition for any infinitely divisible random vector to be so representable is given. The limiting case is discussed as are connections with the multivariate Poisson distribution and stochastic processes.

2. Notation and preliminaries. Let $(\Omega, \mathfrak{B}, \mathcal{P})$ be a probability space; that is, Ω is an abstract point set, $\mathfrak{B}$ a Borel Field of subsets of Ω, and $\mathcal{P}$ a probability measure defined on $\mathfrak{B}$. If $m \geqq 1$ is an integer and $X, Y, Z \cdots$ a set of m-dimensional vectors defined on Ω, we write $X \sim Y$ to signify that the associated probability measures (or distribution functions) of X and Y are identical. Since the relationship indicated by $\sim$ is reflexive, symmetric and transitive, the use of this symbol is in the best of taste and tradition.

We abbreviate the terms cumulative distribution function, characteristic function, random vector, and infinitely divisible by c.d.f., c.f., r.v., and i.d., respectively, and occasionally string some of these together. A bar over a set signifies complementation and the notation R^m is used for m-dimensional Euclidean space.

3. Infinitely divisible vectors. Recall that an r.v. X and likewise its c.d.f., say $F(x_1, x_2, \cdots, x_m)$, and its c.f., say $\varphi(t_1, t_2, \cdots, t_m)$, are called i.d.[2] if for every positive integer n, $X \sim$ sum of n independent (identically distributed) r.v.'s. P. Lévy ([4], p. 220) has given a necessary and sufficient condition (NSC) that X be i.d., viz.,

$$\varphi(t_1, \cdots, t_m) = \exp\left\{ i \sum_{j=1}^{m} \gamma_j t_j - \frac{1}{2} \sum_{j,k=1}^{m} \sigma_{jk} t_j t_k + \int_{R^m} \left[e^{i(t_1 u_1 + \cdots + t_m u_m)} - 1 - \frac{i(t_1 u_1 + \cdots + t_m u_m)}{1 + |u|^2} \right] dN(u_1, \cdots, u_m) \right\}, \tag{1}$$

Received May 28, 1956; revised November 12, 1956.

[1] This work was supported in part by an Office of Naval Research contract at Stanford University.

[2] In many works X is defined to be i.d. if for every positive integer n, $X_n = X_{n1} + X_{n2} + \cdots + X_{nn}$, where $X_{n1}, X_{n2}, \cdots, X_{nn}$ are mutually independent. Such a definition places demands on the basic space Ω. A discussion of this point occurs in Appendix 2 of "Limit Distributions of Sums of Independent Random Variables" by Gnedenko-Kolmogoroff, Addison Wesley. The current definition obviates such questions.

where γ_j are real coefficients $\Sigma = \{\sigma_{ij}\}$ is a positive semidefinite matrix, $|u|$ is the Euclidean length of the row vector $u' = (u_1, u_2, \cdots, u_m)$ and $\mu_N(A) = \int_A dN(u_1, \cdots, u_m)$ is a nonnegative additive set function (not necessarily finite) defined on the Borel sets A of R^m and such that

$$(1') \qquad \int_{S_\epsilon} |u^2| \, dN(u_1, \cdots, u_m) < \infty, \qquad \int_{\bar{S}_\epsilon} dN(u_1, \cdots, u_m) < \infty,$$

(with S_ϵ an m-dimensional sphere of radius $\epsilon > 0$ centered at the origin and $\bar{S}_\epsilon$ its complement in R^m) for arbitrary ϵ.

Let $\varphi(t)$, $N(u)$, and $G(u)$ abbreviate $\varphi(t_1, \cdots, t_m)$, $N(u_1, \cdots, u_m)$ and $G(u_1, \cdots, u_m)$, respectively. Analogous to the one-dimensional case, an alternative and frequently more convenient form of (1) is given by

$$(2) \qquad \varphi(t) = \exp\{i\gamma' t - \tfrac{1}{2} t'\Sigma t\} \cdot \exp\left\{\int_{R^m} \left(e^{it'u} - 1 - \frac{it'u}{1 + |u|^2}\right)\left(\frac{1 + |u|^2}{|u|^2}\right) dG(u)\right\},$$

where t, u, γ are column vectors and

$$\mu_G(A) = \int_A dG(u) = \int_A \frac{|u|^2}{1 + |u|^2} \, dN(u)$$

is, in view of (1'), a finite Lebesgue-Stieltjes measure on the Borel sets of R^m which may be taken to vanish for $A = \{u: |u| = 0\}$.[3] Thus, any i.d.c.f. may be characterized by a triplet $[\gamma, \Sigma, G]$.[4]

The first factor in (2) is obviously the c.f. of a multivariate normal distribution, while the second is generated by the Poisson distribution in a sense which will be made more precise later. Thus, every i.d. vector $X \sim X^{(1)} + X^{(2)}$ where $X^{(1)}$ is multinormal and independent of $X^{(2)}$ which will be said to be "Poisson type." It will be convenient to refer to this as the canonical representation of X. If $X^{(2)} = 0$, X will be called "purely normal" while if $X^{(1)} = 0$, X will be dubbed "purely Poisson type."

If an i.d. vector $X = \begin{pmatrix} U \\ V \end{pmatrix}$ is partitionable into subvectors, one of which (say U) is purely normal and the other V purely Poisson type, then U and V must be independent. For $\begin{pmatrix} U \\ V \end{pmatrix} \sim \begin{pmatrix} U^{(1)} \\ V^{(1)} \end{pmatrix} + \begin{pmatrix} U^{(2)} \\ V^{(2)} \end{pmatrix}$ with $\begin{pmatrix} U^{(1)} \\ V^{(1)} \end{pmatrix}$ purely normal and independent of $\begin{pmatrix} U^{(2)} \\ V^{(2)} \end{pmatrix}$ which is purely Poisson type. But U purely normal implies $U^{(2)} = 0$ and V purely Poisson requires $V^{(1)} = 0$. The assertion follows. This observation may be utilized to construct a non-i.d. vector, all of whose marginal random variables are i.d.

[3] For the most frequently encountered case $m = 1$, this $G(u)$ is not in general identical with that used by the authors of the book mentioned in Footnote 2.

[4] It does not appear to have been remarked (even for the case $m = 1$) that a bounded r.v. is i.d. if and only if it is constant with probability one. This may be argued directly from the definition without resorting to (2).

The fact that for an arbitrary nonzero constant vector

$$c = (c_1, c_2, \cdots, c_m) \quad \text{and c.f.} \quad \varphi_{X_1, X_2, \cdots, X_m}(t_1, t_2, \cdots, t_m),$$
$$\varphi(c_1 t, c_2 t, \cdots, c_m t) = E[\exp\{i(c_1 X_1 + \cdots + c_m X_m)t\}] = \varphi_{c'X}(t)$$

shows immediately that if X is an i.d. vector, every linear combination $c'X$ of its component random variables is i.d. The converse, however, is untrue. That is, it is possible for every linear combination $c'X$ to be i.d. without the vector X being i.d.[5] The Wishart distribution provides an example.

For the c.f. of a so-called Γ-variable is well known to be $[1 - (it/\alpha)]^{-\lambda}$, $\lambda, \alpha > 0$ and is manifestly i.d. Hence, the c.f. $[(1 - it_1)(1 - it_2)]^{-\lambda}$, $\lambda > 0$, of the sum of two independent Γ-variables is clearly i.d., whence by the remarks at the beginning of the preceding paragraph, $[1 - i(c_1 + c_2)t - c_1 c_2 t^2]^{-\lambda}$ is i.d. for arbitrary constants c_1, c_2. But if $Z_j' = (Z_{1j}, Z_{2j})$ are independent normally distributed vectors with mean vector zero and common covariance matrix Σ, then $X' = (X_1, X_2, X_3) = (\sum_1^n Z_{1j}, \sum_1^n Z_{2j}, \sum_1^n Z_{1j}Z_{2j})$ has the Wishart distribution with c.f.

$$|I - \Sigma T|^{-n/2} = [1 - 2i(\sigma_{11}t_1 + \sigma_{22}t_2 + 2\sigma_{12}t_3) + 4(\sigma_{11}\sigma_{22} - \sigma_{12}^2)(t_3^2 - t_1 t_2)]^{-n/2}.$$

Thus, *every* linear combination $b'X$ has the c.f.

$$[1 - 2i(b_1\sigma_{11} + b_2\sigma_{22} + 2b_3\sigma_{12})t + 4(\sigma_{11}\sigma_{22} - \sigma_{12}^2)(b_3^2 - b_1 b_2)t^2]^{-n/2},$$

which is i.d. by the preceding remarks. On the other hand, P. Lévy has shown [5] that the Wishart distribution is not itself i.d. and that for $n = 1$, it is even indecomposable.

If Y is a k-dimensional r.v. whose component random variables are independent and i.d. and A is an arbitrary $m \times k$ matrix of real constants, $X = AY$ is an i.d. r.v. In what sense is the converse true? That is, if X is an i.d. vector when does there exist a constant matrix A and a finite set of *independent* i.d. random variables $Y_1, \cdots, Y_k$ such that $X \sim AY$?

If X is purely normal it is well known that such a representation is always possible. Thus, it suffices to investigate $X^{(2)}$, the Poisson type r.v. in the canonical representation of X. For if $X^{(2)} \sim A_2 Y^{(2)}$ with the k_2 components of $Y^{(2)}$ mutually independent, since $X^{(1)} \sim A_1 Y^{(1)}$ with the k_1 components of $Y^{(1)}$ independent, we will have

$$X \sim X^{(1)} + X^{(2)} \sim (A_1, A_2)\begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = AY,$$

with the $k = k_1 + k_2$ components of Y independent random variables.

An answer to the question posed is given by

THEOREM 1. *A NSC that a Poisson-type r.v. $X \sim AY$ where the components $Y_1, \cdots, Y_k$ of Y are independent non-degenerate i.d. random variables and A is an $m \times k$ matrix of constants, no column of which consists entirely of zeros, is that in (2), μ_σ vanish identically except on k different rays through the origin. Then k is*

[5] It is presumed that the assigned distributions of all linear combinations are compatible with the existence of a joint distribution.

the minimum number of random variables for which a representation $X \sim AY$, with the components $Y_1, \cdots, Y_k$ independent Poisson-type random variables, is possible.

Sufficiency. From the hypothesis of Theorem 1, the c.f. of the i.d. vector X may be supposed characterized by $[\gamma, \Sigma, G]$, with $\gamma' = (0, \cdots, 0)$, $\Sigma = \{0\}$, i.e., $X^{(1)} = (0, \cdots, 0)$. By hypothesis, μ_G assigns positive mass only along k rays, say R_j, whose direction cosine vectors are $r_j' = (r_{1j}, r_{2j}, \cdots, r_{mj})$. Let $G_j(s) = \mu_G(A_j^s)$, where $A_j^s = \{u : u \in R^m, u' = \rho r_j', -\infty < \rho < s\}$ and $h(t'u, |u|)$ denotes the integrand of (2). Then

$$\int_{R^m} h(t'u, |u|)\, dG(u) = \sum_{j=1}^{k} \int_{R^1} h(\rho t' r_j, \rho)\, dG_j(\rho)$$

and

$$\begin{aligned} \varphi_X(t) &= \prod_{j=1}^{k} \exp\left\{ \int_{R^1} \left(e^{i(t'r_j)\rho} - 1 - \frac{i(t'r_j)\rho}{1+\rho^2} \right) \left(\frac{1+\rho^2}{\rho^2} \right) dG_j(\rho) \right\} \\ &= \prod_{j=1}^{k} \varphi_j(t'r_j), \end{aligned} \tag{3}$$

where $\varphi_j(t)$ is a univariate i.d. c.f. characterized by $[0, 0, G_j]$.

Let $Y_1, \cdots, Y_k$ be independent i.d. random variables with Y_j having c.f. $\varphi_j(t)$ as defined in (3) and take A to be the $m \times k$ matrix whose jth column is r_j. Then if $Z = AY$,

$$\begin{aligned} \varphi_{Z_1, \cdots, Z_m}(t_1, \cdots, t_m) &= E\,[\exp\{it'Z\}] = E\,[\exp\{i(t'A)Y\}] \\ &= \prod_{j=1}^{k} E\,[\exp\{i(t'r_j)Y_j\}] = \prod_{j=1}^{k} \varphi_j(t'r_j). \end{aligned} \tag{4}$$

Thus, $X \sim Z = AY$. This representation in terms of the distributions of k independent i.d. random variables (k being the number of rays with positive mass) is unique to within a relabelling of the variables and multiplication of each variable by a nonzero constant. The columns of A must then be adjusted accordingly.

Necessity. Since $X \sim AY$ with the components of Y independent and nondegenerate, the first equality of (3) holds with r_j equal to the jth column of the given matrix $A = \{a_{ij}\}$ divided by the scalar norming factor $(\sum_{i=1}^{m} a_{ij}^2)^{1/2}$. Comparing this with (2), it follows from the uniqueness of the i.d. representation that G and μ_G are as stated in the theorem.

Note that if $k < m$, the mass of the distribution of $X^{(2)}$ is concentrated in a space of lower dimensionality than R^m (i.e., the distribution of $X^{(2)}$ is singular).

A family of distributions $\mathfrak{F} = \{F\}$ has been defined in [7] to be factor-closed if $F = G_1 * G_2$, $F \in \mathfrak{F}$ implies $G_1, G_2 \in \mathfrak{F}$. Then we have as a

COROLLARY. If some component X_i of $X^{(2)}$ has a distribution belonging to a factor-closed family $\mathfrak{F}$, the distributions of $r_{ij}Y_j$ belong to $\mathfrak{F}$, for $j = 1, 2, \cdots, k$.

To avoid trivialities, let all components X_i of X be nondegenerate. Then no row of A is a zero vector. If $X^{(1)} = 0$ and $m = k$, then the components of X (i.e., $X^{(2)}$) are independent if and only if the rows of A may be permuted so as to

form a diagonal matrix. This is palpably sufficient; on the other hand, if A cannot be so juggled, some Y_i has nonzero coefficients in two (necessarily independent) linear forms in the independent random variables Y_j. But this implies [2] that Y_i is normally distributed. The proof of the theorem, however, shows that when $X^{(1)} = 0$, the Y_j are all purely Poisson type, producing a contradiction.

Let α_j' be a k-tuple with 1 in the jth position and zeros elsewhere, $j = 1, \cdots, k$. Then if $k \leqq m$ and all α_j belong to the m-manifold spanned by the m rows of B, no representation of an i.d. vector X in the form BZ with the k components of Z independent but not *all* i.d. random variables is possible. For in such a case $C_j'B = \alpha_j$ has a nonzero solution C_j for all $j = 1, 2, \cdots, k$ whence $C_j'X \sim C_j'BZ = \alpha_j'Z$. But $C_j'X$ and therefore Z_j is i.d. If, e.g., $k > m$ such a representation is not summarily precluded.

It is, in general, untrue that an m-dimensional random vector $Y \sim AX$ where the components of X are independent random variables and the matrix A is $m \times k$. This may be seen with the familiar multinomial distribution.

Example. In r independent repetitions of an experiment, let $Y_1, \cdots, Y_{m+1}$ be the number of occurrences, respectively, of the mutually exclusive and exhaustive outcomes $A_1, \cdots, A_{m+1}$ with (single) trial probabilities $p_1, \cdots, p_{m+1}$, $(\sum_1^{m+1} p_i = 1)$; take $Y' = (Y_1, Y_2, \cdots, Y_m)$ and suppose there exists an integer $k \geqq 1$ and constant vectors $a_j' = (a_{j1}, \cdots, a_{jk})$ such that $Y \sim AX$ with the components of X independent random variables. Then

$$\prod_{j=1}^{k} \varphi_j\left(\sum_{\nu=1}^{m} a_{j\nu} t_\nu\right) = \left[\sum_{\nu=1}^{m} p_\nu (e^{it_\nu} - 1) + 1\right]^r. \tag{5}$$

Setting $t_\mu = t$, $t_\nu = 0$ for $\nu \neq \mu$,

$$\prod_{j=1}^{k} \varphi_j(a_{j\mu} t) = [p_\mu (e^{it} - 1) + 1]^r.$$

Since the classical binomial family is factor-closed [7],

$$\varphi_j(a_{j\mu} t) = e^{ite_j}[p_\mu (e^{it} - 1) + 1]^{r_j} \text{ with } 0 < r_j \leqq r,$$
$$\sum r_j = r, \ \sum c_j = 0,$$

or

$$\varphi_j(t) = e^{ite_j/a_{j\mu}}[p_\mu (e^{it/a_{j\mu}} - 1) + 1]^{r_j}, \qquad j = 1, 2, \cdots, k.$$

Since the left-hand side is independent of μ, so is the right-hand side, whence $p_\mu = p$, $a_{j\mu} = a_j$, $\mu = 1, 2, \cdots, m$. Thus, if the multinomial probabilities are not identical, (5) cannot hold. However, even if $p_i \equiv p$, (5) would imply

$$\prod_{j=1}^{k} \varphi_j\left(\sum_{\nu=1}^{m} a_{j\nu} t_\nu\right) = \prod_{j=1}^{k} \left[p\left(\exp\left[i \sum_{\nu=1}^{m} t_\nu\right] - 1\right) + 1\right]^{r_j}$$
$$\equiv \left[\sum_{\nu=1}^{m} p(e^{it_\nu} - 1) + 1\right],$$

which is impossible since the middle expression is a function of $\sum_{i=1}^{m} t_i$ only (and hence a degenerate multivariate distribution) whereas the right-hand side is not.

The following theorem covers the case that the measure μ_G is not necessarily wholly concentrated on a finite number of rays through the origin.

THEOREM 2. *If X is an i.d. random vector, then there exists a sequence of vectors $\{Y_n\}$ consisting of independent i.d. components, and a sequence of finite matrices $\{A_n\}$ such that the distribution of A_nY_n converges to the distribution of X as $n \to \infty$. The components of Y_n can each be taken to be of the form $(Y - b)$, where Y is a Poisson random variable if X is purely Poisson type.*

PROOF. As earlier, we may suppose $X^{(1)}$ is zero. Let

$$h(u) = \left(e^{it'u} - 1 - \frac{it'u}{1 + |u|^2}\right)\left(\frac{1 + |u|^2}{|u|^2}\right).$$

There is a double sequence of positive constants

$$\lambda_{n,1}, \cdots, \lambda_{n,k(n)}, \qquad n = 1, 2, \cdots$$

and a double sequence of m-tuples,

$$u_{n,1}, \cdots, u_{n,k(n)}, \qquad n = 1, 2, \cdots$$

$$u'_{n,i} = (u^{(1)}_{n,i}, \cdots, u^{(m)}_{n,i}), \qquad i = 1, 2, \cdots, k(n)$$

such that

$$\sum_{i=1}^{k(n)} \lambda_{n,i} h(u_{n,i}) \to \int_{R^m} h(u)\, dG(u), \tag{6}$$

as $n \to \infty$. Now $\lambda_{n,i}h(u_{n,i})$ is the log of the c.f. of a random vector

$$([Y_{ni} - b_{ni}]u^{(1)}_{ni}, \cdots, [Y_{ni} - b_{ni}]u^{(m)}_{ni}), \tag{7}$$

where the b_{ni} are appropriately chosen constants, and Y_{ni} is a Poisson random variable with parameter λ_{ni}. Hence, the left-hand side of (6) is the log of the c.f. of a sum of $k(n)$ vectors of the form (7), where $Y_{n1}, \cdots, Y_{nk(n)}$ are mutually independent. In other words, the left-hand side of (6) is the log of the c.f. of the vector

$$\begin{pmatrix} u^{(1)}_{n1} & & u^{(1)}_{n,k(n)} \\ \cdot & & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & & \cdot \\ u^{(m)}_{n1} & & u^{(m)}_{n,k(n)} \end{pmatrix} \cdot \begin{pmatrix} \bar{Y}_{n1} \\ \cdot \\ \cdot \\ \cdot \\ \bar{Y}_{nk(n)} \end{pmatrix},$$

where $\bar{Y}_{ni} = [Y_{ni} - b_{ni}]$, $i = 1, \cdots, k(n)$ are mutually independent i.d. random variables. This completes the proof.

4. Multivariate Poisson distribution. Let V denote the set of $2^m - 1$ vertices (excluding the origin) of the unit cube in the first orthant of R^m and lying along

the m-axes; let V_j signify the vertex with one as the jth coordinate and zeros for the others, $j = 1, 2, \cdots, m$; let V_{ij}, $i < j$, represent the vertex with one's for the ith and jth coordinates but zeros for the remaining; $\cdots$; finally, let $V_{1,2,\cdots,m}$ denote the vertex $(1, 1, \cdots, 1)$.

In (2), define $G(u)$ by $\mu_G(V_j) = a_j \geqq 0, j = 1, \cdots, m$, $\mu_G(V_{ij}) = a_{ij} \geqq 0$, $i < j; \cdots, \mu_G(V_{1,2,\cdots,m}) = a_{1,2,\cdots,m} \geqq 0$, and for any Borel set B of R^m, $\mu_G(B) = \mu_G(VB)$ where the measure of the empty set is zero. Then if $z_j = e^{it_j}$, (2) becomes

$$(8) \qquad \varphi(t) = \exp\left\{\sum_{j=1}^{m} a_j z_j + \sum_{i<j} a_{ij} z_i z_j + \cdots + a_{1,2,\cdots,m} \prod_{j=1}^{m} z_j - A_m\right\},$$

where A_m is a constant such that $\varphi(0, 0, \cdots, 0) = 1$. The c.f. in (8) is that of the multivariate Poisson discussed in [6]. Since $G(u)$ is of the form prescribed by Theorem 1, with $k = 2^m - 1$, it follows from this theorem (supposing the constants $a_j, a_{ij}, \cdots, a_{12,\cdots,m}$ strictly positive) that there are $2^m - 1$ random variables Y_j and a constant matrix A such that $X \sim AY$. The matrix A may be chosen so that its $2^m - 1$ columns are the vectors (vertices) of V. By the corollary to Theorem 1, the Y_j are also Poisson distributed with parameters $a_1, a_2, \cdots, a_m; a_{12}, a_{13}, \cdots, a_{m-1,m}; \cdots; a_{1,2,\cdots,m}$. Since the classical Poisson distribution is not invariant under scale change, the matrix A is uniquely determined to within a permutation of its columns by the stipulation that the Y_j be independent Poisson random variables.

Furthermore, the multivariate Poisson distributions specified in (8) are the only i.d. distributions which are marginally Poisson. For, under this last proviso, $G(u)$ in (2) must be such that the projection of μ_G on the jth coordinate axis concentrates all mass at the point $(0, 0, \cdots, 0, 1, 0, \cdots, 0)$. This, in turn, requires that μ_G be as defined in the previous paragraph.

More generally, let $\mathfrak{F} = \{F(x; b_1, \cdots, b_r; c_1, \cdots, c_r\}$ be a family of univariate i.d. distributions whose c.f.'s are characterized by $[0, 0, G]$ with μ_G a discrete measure assigning mass $c_h > 0$ to $u = b_h \neq 0, h = 1, 2, \cdots, r$. Let X be an i.d. vector with the prescribed marginal distribution $F_{X_i}(x) = F(x; b_1, \cdots, b_r; c_1^i, \cdots, c_r^i)$, $i = 1, 2, \cdots, m$. Then, as earlier, there is a unique family of i.d. distributions for X having the stated marginals. Its c.f.'s are characterized by $[\gamma, \Sigma, G_d]$ where $\gamma' = (0, 0, \cdots, 0)$, $\Sigma = \{0\}$, and μ_{G_d} is a discrete measure assigning mass $d_j \geqq 0$ to the $(r + 1)^m - 1$ points $(u_1, u_2, \cdots, u_m)$ where $u_i = b_i$ or 0 (but u_i not all zero). Here the independent random variables may be taken to have the classical Poisson distribution and

$$X \sim DY = \begin{pmatrix} \sum_{j=1}^{k} d_{1j} Y_j \\ \cdot \\ \cdot \\ \cdot \\ \sum_{1}^{k} d_{mj} Y_j \end{pmatrix} = \sum_{j=1}^{k} \begin{pmatrix} d_{1j} Y_j \\ \cdot \\ \cdot \\ \cdot \\ d_{mj} Y_j \end{pmatrix}.$$

It is degenerate vectors of this form based on a single Poisson random variable Y_j rather than nondegenerate vectors having the most general multivariate Poisson distribution that spawn i.d. vectors.

It was pointed out above that if X is a multivariate i.d. vector, all of whose components have Poisson distributions marginally, then X must have a multivariate distribution specified by (8) and $X \sim AY$ when A is a finite matrix and Y is a vector of independent Poisson random variables. The purpose of the next remarks is to indicate that in general a comparable situation does not prevail. For example, suppose that U_1, U_2, U_3 are independent gamma variables whose c.f.'s are all $(1-it)^{-\lambda}$, $(\lambda > 0)$. Then

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 101 \\ 011 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}$$

is an i.d. vector with c.f. $\{(1 - it_1)(1 - it_2)[1 - i(t_1 + t_2)]\}^{-\lambda}$ whose marginals X_1, X_2 are gamma variables. On the other hand, in [3] it is shown that if $|\rho| < 1$. then

$$[(1 - it_1)(1 - it_2) + \rho^2 t_1 t_2]^{-2\lambda} \tag{9}$$

is a c.f. for all $\lambda > 0$ (and hence i.d.) and its marginals clearly have the same distribution as do X_1 and X_2. Thus there is no unique i.d. family having gamma marginals. Suppose $\rho \neq 0$ to avoid the trivial case of independence; then it is easy to verify that (9) cannot be the c.f. of a finite linear combination of independent gamma variables.

5. Connection with stochastic processes. It is a familiar fact that in the one-dimensional case the theory of i.d. random variables has a close connection with the theory of stochastic processes with independent increments. The analogue for multivariate i.d. vectors should be apparent, but it may be worth making some of the facts explicit.

Let U be a random vector whose values are the vectors of the set V defined at the beginning of Section 4. Denote these values by $u_1, \cdots, u_k$ and let their corresponding probabilities be $p_1, \cdots, p_k$, where $k = 2^m - 1$. Let $U_1, U_2, \cdots$, be an infinite sequence of independent random vectors, each distributed as U. Let $X'(t)$, $(t \geqq 0, X'(0) = 0)$, be a Poisson process with stationary, independent increments. It is well known that waiting times for "jumps" in $X'(t)$ are independent, identically distributed exponential random variables. That is, an equivalent way of defining this process is in terms of an infinite sequence of independent, identically distributed random variables $W_1, W_2, \cdots$, such that $P(W_i < w) = \lambda \int_0^w e^{-\lambda y}\, dy$ for $w > 0$ and zero otherwise $(\lambda > 0)$ as follows:

$$X'(0) = 0,\ 0 \leqq t \leqq W_1,$$

$$X'(t) = 1,\ W_1 < t \leqq W_1 + W_2,$$

$$X'(t) = 2, W_1 + W_2 < t \leqq W_1 + W_2 + W_3,$$

etc. Analogously, we can now define a *multivariate* Poisson process $X(t)$ as follows:

$$X(0) = 0, \text{ (zero } m\text{-vector)}, 0 \leqq t \leqq W_1,$$

$$X(t) = U_1, W_1 < t \leqq W_1 + W_2,$$

$$X(t) = U_1 + U_2, W_1 + W_2 < t \leqq W_1 + W_2 + W_3,$$

etc. Making use of the well known fact that the conditional distribution of $W_1, \cdots, W_r$ given that $X'(t) = r$ is that of the ordered values of r independent random variables, each uniformly distributed in $(0, t)$, it is easy to compute that the c.f. of $X(t)$ is

$$\sum_{j=0}^{\infty} [C(\theta)]^j \frac{(\lambda t)^j e^{-\lambda t}}{j!} = \exp\left\{\lambda t \sum_{j=1}^{k} (e^{i\theta' u_j} - 1)\, p_j\right\},$$

where

$$C(\theta) = C(\theta_1, \cdots, \theta_m) = \sum_{j=1}^{k} e^{i\theta' u_j} p_j$$

is the c.f. of the random vector U and $u_1, \cdots, u_k$ is the set of the k possible values of U. Making use of the material in Section 4, we see that we can choose the p_j's and λ so that $X(t)$ has any prescribed i.d. multivariate Poisson distribution. We remark also that $X(t)$ has independent, stationary increments for exactly the same reasons that $X'(t)$ does.

Consider now the somewhat more general case in which $U_1, U_2, \cdots$ are independent, identically distributed random vectors (m-tuples) having an *arbitrary* distribution with c.f.

$$C(\theta) = \int_{R_m} e^{i\theta' u}\, dF(u),$$

where F is the distribution function of U_1. If we define $X(t)$ as above but in terms of these more general U_i's, then the c.f. of $X(t)$ is

$$\exp\left\{\lambda t \int_{R_m} (e^{i\theta' u} - 1)\, dF(u)\right\}.$$

We recognize this to be a multivariate i.d. c.f. either from the Lévy form or from the fact that $X(t)$ has independent increments. We cannot obtain the most general multivariate i.d. c.f. in this way. On the other hand, we can find a sequence of constant vectors $a_1, a_2, \cdots$ and scalars $b_1, b_2, \cdots$ and distribution functions $F_1, F_2, \cdots$ such that if $X_n(t)$ is determined by F_n as above, then as $n \to \infty$,

$$(X_n(t) - a_n)/b_n$$

converges in law to an arbitrary Poisson i.d. vector. Thus, the most general Poisson-type i.d. vector can be approximately obtained in terms of a Poisson-like stochastic process with independent exponential waiting times between "jumps" and whose "jumps" are random vectors.

REFERENCES

[1] SALOMON BOCHNER, *Harmonic Analysis and the Theory of Probability*, University of California Press, Berkeley and Los Angeles, 1955.

[2] G. DARMOIS, "Sur une propriété caractéristique de la loi de probabilité de Laplace," *C. R. Acad. Sci. Paris*, Vol. 232 (1951), pp. 1999–2000.

[3] A. S. KRISHNAMOORTHY AND M. PARTHASARATHY, "A multivariate gamma-type distribution," *Ann. Math. Stat.*, Vol. 22 (1951) pp. 549–557.

[4] PAUL LÉVY, *Théorie de l'Addition des Variable Aléatoires*, Gauthier-Villars, Paris, 1937, pp. 214–221.

[5] PAUL LÉVY, "Arithmetical character of Wishart distribution," *Proc. Cambridge Philos. Soc.*, Vol. 44 (1948), pp. 295–297.

[6] HENRY TEICHER, "On the multivariate Poisson distribution," *Skand. Aktuarietids.* (1954), pp. 1–9.

[7] HENRY TEICHER, "On the factorization of distributions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 769–774.

A MOVING SINGLE SERVER PROBLEM

By B. McMillan and J. Riordan

Bell Telephone Laboratories, Incorp., New York

1. Introduction. An assembly line moving with uniform speed has items for service spaced along it. The single server available moves with the line while serving and against it with infinite velocity while transferring service to the next item in line. The line has a barrier in which the server may be said to be "absorbed" in the sense that service is disabled if the server moves into the barrier. The problem solved here is the following: given that a server with exponentially distributed service time starts service on the first item when it is T time units away from the barrier, what is the probability $p(k, T)$ that it completes k items of service before absorption? This is the same as determining the generating function

$$P(x, T) = \sum_{k=0}^{\infty} p(k, T)x^k. \tag{1}$$

The referee has pointed out to us an identification of this problem with that of finding the number of units of service in a busy period for the usual (stationary) single server. This may be seen as follows.

Take $\tau(t)$ as the distance from the barrier at time t, so that $\tau(0) = T$. Take the spacing between items as an independent random variable with distribution function $B(t)$. Then the graph of $\tau(t)$ as in Fig. 1 consists of lines of unit slope interrupted by jumps having the distribution $B(t)$ and occurring at t-epochs determined by the exponential distribution of service time. The graph ends when $\tau(t) = 0$ for the first time, when service is disabled.

Now consider the queueing system with a single server, Poisson arrivals, and distribution of service times $B(t)$. Take $\tau(t)$ as the waiting time of a *virtual* arrival at time t. Then the graph of $\tau(t)$ for a single busy period of the server is exactly as in Fig. 1 if the first customer served has a service time which is *given* to be T.

Note that one problem is turned into the other by interchanging service and arrival variables.

Busy periods were first considered by E. Borel [2] for the case of constant service time and with main interest in the number served, exactly as here, but with the first customer's service time the same constant as all others. Turning to the length of the busy period, D. G. Kendall [4] generalized Borel's result to arbitrary service time distribution by transforming it into a question concerning a branching process. Kendall's functional equation was carefully derived by L. Takacs [7], who also obtained a similar equation for the generating function for the number served in a busy period (with no condition on the first customer) for

Received July 5, 1956; revised October 5, 1956.

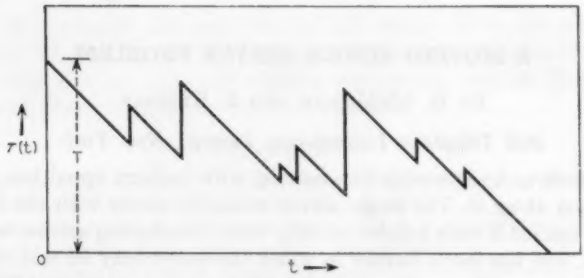


FIG. 1. Sample behavior of random variable $\tau(t)$
$\tau(t)$ = distance from barrier at time t (moving server)
= waiting time of a virtual arrival at time t (queueing system)

arbitrary distribution of service time. All of these are under the usual assumption of Poisson arrivals.

Takacs' result (l.c. Theorem 7, p. 120) in present notation is as follows:

THEOREM (Takacs). *If the generating function of number served in a busy period is*

$$F(x) = x \int_0^\infty P(x, T)\, dB(T), \tag{2}$$

and if arrivals are Poisson with average α in unit time, then

$$F(x) = x \int_0^\infty \exp\,[-\alpha t(1 - F(x))]\, dB(t). \tag{3}$$

This suggests that the conditional generating function $P(x, T)$ satisfies

$$P(x, T) = \exp\left\{-\alpha T\left[1 - x \int_0^\infty P(x, T)\, dB(T)\right]\right\}, \tag{4}$$

which deserves an independent derivation. It is clear that it is not a simple consequence of (3), since in the case of constant service time equal to ϵ

$$F(x) = xP(x, \epsilon)$$

and cannot possibly determine $P(x, T)$ for arbitrary T. Nevertheless Eq. (4) is correct.

Because of this, we retain our original derivation of $P(x, T)$ which is limited to the two extreme cases (of most interest to us), namely (i) constant spacing

$$B(t) = 0, \quad t \leqq \epsilon, \tag{5}$$
$$= 1, \quad t > \epsilon,$$

and (ii) random spacing

$$B(t) = 0, \quad t \leqq 0, \tag{6}$$

$$= 1 - e^{-\beta t}, \qquad t > 0,$$

for both of which we take the service distribution as

$$A(t) = 0, \qquad t \leqq 0, \tag{7}$$
$$= 1 - e^{-\alpha t}, \qquad t > 0.$$

The average service time is $1/\alpha$.

We show that (4) is true in both cases and obtain explicit expressions for the probabilities $p(k, T)$ and for their moments.

2. Uniform spacing. The probability $p\ (0, T)$, that service on the first item begun T time units away from the barrier is not completed before absorption, is the probability that the service time is greater than T; hence

$$p(0, T) = 1 - A(T) = e^{-\alpha T}. \tag{8}$$

For the other probabilities $p(k, T)$, $k = 1, 2, \cdots$, a recurrence may be found as follows. Suppose service on the first item is completed in the interval $t, t + dt$; then service is begun on the second item at a point $t + \epsilon$ time units away from the barrier, and it follows at once that

$$\begin{aligned} p(k, T) &= \int_0^T p(k - 1, t + \epsilon)\, dA(T - t), \\ &= \int_0^T p(k - 1, t + \epsilon)\alpha e^{-\alpha(T-t)}\, dt. \end{aligned} \tag{9}$$

Then, using Eq. (1), the generating function $P(x, T)$ must satisfy

$$P(x, T) = e^{-\alpha T} + x \int_0^T P(x, t + \epsilon)\alpha e^{-\alpha(T-t)}\, dt. \tag{10}$$

Suppose that this has a solution $e^{-\lambda T}$, $\lambda \equiv \lambda(x; \alpha, \epsilon)$; then (10) shows that

$$e^{-\lambda T} - e^{-\alpha T} = \alpha x e^{-\lambda\epsilon}(e^{-\lambda T} - e^{-\alpha T})/(\alpha - \lambda), \tag{11}$$

or

$$\alpha - \lambda = \alpha x e^{-\lambda\epsilon}. \tag{11a}$$

But this is what (4) becomes when $B(t)$ is given by (5) and $P(x, T) = e^{-\lambda T}$.

Notice that for $x = 0$, $\lambda = \alpha$, as is required by (8). Note also that all probabilities $p(k, T)$, $k = 1, 2, \cdots$ are uniquely determined by (8) and (9), and that $P(x, T)$ is an analytic function for $x < 1$. To determine it rewrite (11a) in the form

$$(\alpha\epsilon - \lambda\epsilon)e^{-(\alpha\epsilon-\lambda\epsilon)} = x\alpha\epsilon e^{-\alpha\epsilon}$$

or, what is the same thing,

$$ze^{-z} = w, \qquad z = \alpha\epsilon - \lambda\epsilon, \qquad w = x\alpha\epsilon e^{-\alpha\epsilon}. \tag{12}$$

This is an equation familiar in Lagrange series expansions and in fact the expansion for $\exp(zT/\epsilon) = \exp(\alpha T - \lambda T)$ is given by Pólya and Szegö [5] (III Abschnitt, p. 210) in the form

$$\exp(\alpha T - \lambda T) = 1 + \sum_{k=1}^{\infty} \frac{(T/\epsilon)(T/\epsilon + k)^{k-1}}{k!} w^k, \qquad w < e^{-1}$$

or

$$\exp -\lambda T = e^{-\alpha T} + \sum_{k=1}^{\infty} \frac{T(T + k\epsilon)^{k-1}}{k!} (\alpha\epsilon^{-\alpha\epsilon})^k e^{-\alpha T} x^k. \tag{13}$$

Hence

$$p(k, T) = \frac{T(T + k\epsilon)^{k-1}}{k!} (\alpha\epsilon^{-\alpha\epsilon})^k \epsilon^{-\alpha T}, \tag{14}$$

a result which may also be obtained from (9) and mathematical induction.

For the probability $P(1, T)$ of absorption, (12) becomes

$$ze^{-z} = \alpha\epsilon e^{-\alpha\epsilon}. \tag{15}$$

The function $y(x) = xe^{-x}$ of the real variable x is zero for x zero, increases to a maximum at $x = 1$ and decreases monotonically to zero; hence the equation $a - xe^{-x} = 0$ has two real roots for $a < e^{-1}$ and in the present instance, Eq. (11), because probabilities are in question, the smaller is the proper one. For $\alpha\epsilon < 1$, this root is $\alpha\epsilon$ itself, otherwise it is denoted by z_0. Hence

$$\begin{aligned} P(1, T) &= 1, && \alpha\epsilon \leqq 1, \\ &= \exp[-(\alpha - z_0/\epsilon)T], && \alpha\epsilon > 1. \end{aligned} \tag{16}$$

It is interesting to notice that the first of these may be verified as follows. Rewrite (14) as

$$p(k, T) = e^{-\alpha T}(\alpha\epsilon^{-\alpha\epsilon})^k \left[\frac{(T + k\epsilon)^k}{k!} - \epsilon \frac{(T + k\epsilon)^{k-1}}{(k-1)!}\right]. \tag{14a}$$

Then, by a result given by Jensen [3], namely

$$\sum_0^{\infty} e^{-(a+kx)} \frac{(a + kx)^k}{k!} = \frac{1}{1 - x}, \qquad |x| < 1$$

and (14a), it follows that

$$P(1, T) = \frac{1}{1 - \alpha\epsilon} - \frac{\alpha\epsilon}{1 - \alpha\epsilon} = 1, \qquad \alpha\epsilon < 1.$$

Jensen's result may also be used with (14) to show that

$$M(T) = \sum kp(k, T) = \alpha T(1 - \alpha\epsilon)^{-1}, \qquad \alpha\epsilon < 1. \tag{17}$$

For higher moments, two courses are open. First, since

$$P(1 + x, T) = \sum_0^{\infty} x^k M_{(k)}(T)/k! = M(x, T) \tag{18}$$

with $M_{(k)}(T)$ the kth factorial moment, it follows from (10) that

$$M(x, T) = e^{-\alpha T} + \alpha(1 + x)\int_0^T M(x, t + \epsilon)e^{-\alpha(T-t)}\,dt. \tag{19}$$

By differentiation

$$\partial M(x, T)/\partial T = \alpha[M(x, T + \epsilon) - M(x, T) + xM(x, T + \epsilon)]; \tag{20}$$

hence, equating powers of x, with a prime denoting a derivative,

$$M'_{(k)}(T) = \alpha M_{(k)}(T + \epsilon) - \alpha M_{(k)}(T) + \alpha k M_{(k-1)}(T + \epsilon), \tag{21}$$

a differential recurrence relation which may be solved step by step, and which is satisfied by $M_{(1)}(T) = M(T)$, where $M(T)$ is given by (17). The next case is

$$M'_{(2)}(T) = \alpha M_{(2)}(T + \epsilon) - \alpha M_{(2)}(T) + 2\alpha^2(T + \epsilon)(1 - \alpha\epsilon)^{-1}$$

and it turns out that

$$M_{(2)}(T) = \alpha T(\alpha\epsilon)(2 - \alpha\epsilon)(1 - \alpha\epsilon)^{-2} + M^2(T).$$

Second, from (12) by Lagrangian inversion (cf [5], l.c. 209)

$$z = w + \frac{2w^2}{2!} + \cdots + (n)^{n-1}\frac{w^n}{n!} + \cdots$$

and

$$\begin{aligned} \exp Tz/\epsilon &= \exp(\alpha T - \lambda T) \\ &= \exp(xuT + 2\epsilon T(xu)^2/2! + \cdots + (n\epsilon)^{n-1}T(xu)^n/n! + \cdots) \\ &= \sum (xu)^n Y_n(y_1, y_2, \cdots, y_n)/n! \\ &= \exp xuY, \quad \text{symbolically,} \qquad \alpha\epsilon < 1, \end{aligned} \tag{22}$$

with $u = \alpha e^{-\alpha\epsilon}$, $Y_n(y_1, y_2, \cdots, y_n)$ a multivariable polynomial introduced by Bell [1], $y_n = (n\epsilon)^{n-1}T$, and in the symbolic abbreviation the usual convention: $Y^k \equiv Y_k(y_1, y_2, \cdots, y_k)$ is followed. (The relation used in the second and third lines of (18) may be regarded as a definition of the Y polynomials).

Then

$$M(x, T) = \exp[(1 + x)uY - \alpha T], \quad \text{symbolically,}$$

and again for $\alpha\epsilon < 1$

$$\begin{aligned} M_{(k)}(T) &= e^{-\alpha T}u^k D^k \exp uY, \qquad D = d/du, \\ &= e^{-\alpha T}u^k D^k \exp \alpha T \\ &= Y_k(Tu\alpha_1, Tu^2\alpha_2, \cdots, Tu^k\alpha_k), \qquad \alpha_k = D^k\alpha, \end{aligned} \tag{23}$$

the second line following from $M(0, T) = 1$, the third from the development in [6].

The derivatives α_k are readily calculated; indeed, from the initial values

$$u\alpha_1 = \alpha(1 - \alpha\epsilon)^{-1}, \qquad u^2\alpha_2 = \alpha(\alpha\epsilon)(2 - \alpha\epsilon)(1 - \alpha\epsilon)^{-3}$$

and mathematical induction, it is found that

$$u^k\alpha_k = \alpha(1-v)^{1-2k}q_k(v), \qquad v = \alpha\epsilon \tag{24}$$

with

$$q_{k+1}(v) = [1 - k + (4k-2)v - kv^2]q_k(v) + v(1-v)q'_k(v),$$

the prime indicating a derivative.

It may be noticed that the variance of the number served is given by

$$\begin{aligned}\text{var} &= M_{(2)}(T) + M(T) - M^2(T)\\ &= \alpha T(1-\alpha\epsilon)^{-3}.\end{aligned}$$

3. Random spacing. As before $p(0, T) = e^{-\alpha T}$, and the other probabilities are obtained by a recurrence derived as follows. Suppose service on the second item is begun when it is in the interval $(S, S + dS)$ in time units away from the barrier; the probability of this event is, with $a = \alpha\beta/(\alpha+\beta)$,

$$\beta\, dS \int_0^T e^{-\beta(S+t-T)}\alpha e^{-\alpha t}\, dt = a(e^{\beta T} - e^{-\alpha T})e^{-\beta S}\, dS, \qquad S > T,$$

and

$$\beta\, dS \int_{T-S}^T e^{-\beta(S+t-T)}\alpha e^{-\alpha t}\, dt = ae^{-\alpha T}(e^{\alpha S} - e^{-\beta S})\, dS, \qquad S < T.$$

Hence, just as with (9)

$$\begin{aligned}p(k, T) = {} & ae^{-\alpha T}\int_0^T (e^{\alpha S} - e^{-\beta S})p(k-1, S)\, dS\\ & + a(e^{\beta T} - e^{-\alpha T})\int_T^\infty e^{-\beta S}p(k-1, S)\, dS, \qquad k > 0.\end{aligned} \tag{25}$$

It may be noticed for verifications that

$$\begin{aligned}p(1, T) &= aTe^{-\alpha T},\\ p(2, T) &= a^2Te^{-\alpha T}(\alpha+\beta)^{-1} + \alpha^2T^2e^{-\alpha T}/2!.\end{aligned}$$

The probability generating function $P(x, T)$, defined by (1), has the recurrence

$$\begin{aligned}P(x, T) = {} & e^{-\alpha T} + axe^{-\alpha T}\int_0^T (e^{\alpha S} - e^{-\beta S})P(x, S)\, dS\\ & + ax(e^{\beta T} - e^{-\alpha T})\int_T^\infty e^{-\beta S}P(x, S)\, dS.\end{aligned} \tag{26}$$

Trying an exponential solution

$$P(x, T) = e^{-\lambda T}, \qquad \lambda \equiv \lambda(x; \alpha, \beta)$$

leads to the conditional (quadratic) equation

$$(\alpha - \lambda)(\beta + \lambda) = \alpha\beta x, \tag{27}$$

which again agrees with (4) when $B(t)$ is given by (6) and $P(x, T) = \exp -\lambda T$ as above. The solution of (27) is

$$2\lambda = \alpha - \beta + [(\alpha + \beta)^2 - 4\alpha\beta x]^{1/2}.$$

The positive sign must be chosen since it leads to $\lambda(0) = \alpha$ and $P(x, T) \leqq 1$ for $x \leqq 1$. Hence

$$P(x, T) = \exp -\frac{T}{2}[\alpha - \beta + \sqrt{(\alpha + \beta)^2 - 4\alpha\beta x}]. \tag{28}$$

It follows at once (taking the positive square root) that

$$\begin{aligned} P(1, T) &= 1, && \alpha \leqq \beta, \\ &= e^{-(\alpha-\beta)T}, && \alpha \geqq \beta. \end{aligned} \tag{29}$$

The probabilities $p(k, T)$ can be obtained easily from the generating function by noting that its second derivative may be written as

$$[(\alpha + \beta)^2 - 4\alpha\beta x]P''(x, T) = 2\alpha\beta P'(x, T) + (\alpha\beta T)^2 P(x, T). \tag{30}$$

From this follows the recurrence

$$\begin{aligned} &(k + 2)(k + 1)p(k + 2, T) \\ &\quad = (2k + 2)(2k + 1)(a^2/\alpha\beta)p(k + 1, T) + a^2T^2p(k, T). \end{aligned} \tag{31}$$

For an explicit expression, write

$$p(k, T) = e^{-aT}\sum_{j=0}^{k-1} A_{kj}\frac{(aT)^{k-j}}{(k - j)!}b^j, \qquad b = a^2/\alpha\beta;$$

then the numbers A_{kj} are determined by the generating function recurrence

$$\begin{aligned} (1 - x)A_k(x) &= (1 - x)\sum A_{kj}x^j \\ &= A_{k-1}(x) - \frac{1}{k}\binom{2k - 2}{k - 1}x^k. \end{aligned}$$

Similarly factorial moments are determined from the following relation for the derivatives of $M(x, T) = P(1 + x, T)$:

$$[(\alpha - \beta)^2 - 4\alpha\beta x]M''(x, T) = 2\alpha\beta M'(x, T) + (\alpha\beta T)^2 M(x, T), \tag{32}$$

which leads to the recurrence

$$M_{(k+2)}(T) = (4k + 2)[\alpha\beta/(\alpha - \beta)^2]M_{(k+1)}(T) + [(\alpha\beta T)^2/(\alpha - \beta)^2]M_{(k)}(T). \tag{33}$$

Hence

$$M_{(k)}(T) = k!\sum_{j=0}^{k-1} A_{kj}\frac{(cT)^{k-j}}{(k - j)!}d^j$$

with the numbers A_{kj} as above, and $c = \alpha\beta/(\beta - \alpha)$, $d = \alpha\beta/(\beta - \alpha)^2$.

The mean and variance of the number served are

$$M(T) = \frac{\alpha\beta T}{\beta - \alpha} = \frac{\alpha T}{1 - \alpha/\beta},$$

$$\text{var } (T) = \frac{\alpha\beta(\beta^2 - \alpha^2)T}{(\beta - \alpha)^3} = \frac{\alpha T(1 - (\alpha/\beta)^2)}{(1 - \alpha/\beta)^3}.$$

Note the similarity to the corresponding results for uniform spacing.

Thanks are due to E. N. Gilbert for a thoroughgoing review of an earlier draft, and to the referee for his stimulating identification of two apparently distinct problems.

REFERENCES

[1] E. T. Bell, "Exponential polynomials," *Ann. Math.*, Vol. 35 (1934), pp. 258–277.

[2] E. Borel, "Sur l'emploi du theoreme de Bernoulli pour faciliter le calcul d'une infinite de coefficients. Application au probleme de l'attente a un guichet," *C. R. Acad. Sci. Paris*, Vol. 214 (1942), pp. 452–456.

[3] J. L. W. V. Jensen, "Sur une identite d'Abel et sur d'autres formules analogues," *Acta Math.*, Vol. 26 (1902), pp. 307–318.

[4] D. G. Kendall, "Some problems in the theory of queues," *J. Roy Stat. Soc.*, Ser. B., Vol. 13 (1951), pp. 151–173.

[5] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, I., New York, 1945.

[6] J. Riordan, "Derivatives of composite functions," *Bull. Amer. Math. Soc.*, Vol. 52 (1946), pp. 664–667.

[7] L. Takacs, "Investigation of waiting time problems by reduction to Markov processes," *Acta Math. Acad. Sci. Hungaricae*, Vol. 6 (1955), pp. 101–129.

SOME FURTHER METHODS OF CONSTRUCTING REGULAR GROUP DIVISIBLE INCOMPLETE BLOCK DESIGNS

By G. H. Freeman

East Malling Research Station, Maidstone, England

1. Summary. Some further methods are given for the construction of regular group divisible incomplete block designs, and designs derivable by these methods are tabulated. The methods are (i) designs containing complete and incomplete groups; (ii) designs with groups arranged in sets; and (iii) designs derivable by addition. In the first two of these methods, which are related, it may be possible to avoid having to take all the blocks that the full procedure would indicate.

2. Introduction. An incomplete block design with r replicates of v treatments on b blocks with k plots in each is said to be group divisible if it contains m groups of n treatments each, where $mn = v$, in such a way that treatments in the same group concur in λ_1 blocks and treatments in different groups concur in λ_2 blocks, where $\lambda_1 \neq \lambda_2$. For all group divisible designs the following relationships hold: $bk = vr$, $\lambda_1(n-1) + \lambda_2 n(m-1) = r(k-1)$, $r \geqq \lambda_1$, $rk \geqq \lambda_2 v$. Further, the efficiency factors are as follows:

$$\text{within groups,} \quad E_1 = \frac{n\{\lambda_1 + (m-1)\lambda_2\}}{rk} = 1 - \frac{r - \lambda_1}{rk};$$

$$\text{between groups,} \quad E_2 = \frac{mn^2\lambda_2\{\lambda_1 + (m-1)\lambda_2\}}{rk\{\lambda_1 + (mn-1)\lambda_2\}} = \frac{\lambda_2 v}{\lambda_1 + \lambda_2(v-1)} E_1.$$

Group divisible designs have been classified into three types by Bose and Connor [2]: (i) singular designs for which $r = \lambda_1$; (ii) semi-regular designs for which $r > \lambda_1$, $rk = \lambda_2 v$; (iii) regular designs for which $r > \lambda_1$, $rk > \lambda_2 v$. It is the purpose of this paper to present some unpublished methods for the construction of regular group divisible designs, and to give examples of the designs obtained by these methods.

Methods for the construction of group divisible incomplete block designs have been given by Bose, Shrikhande, and Bhattacharya [3], and tables of such designs, *inter alia*, have been prepared by Bose, Clatworthy, and Shrikhande [1]. The designs derived here are of the following kinds: (i) designs containing complete and incomplete groups; (ii) designs with groups arranged in sets; (iii) designs derivable by addition. The first two of these are related in that designs of a very general nature which are generalizations of the first kind are also, in one sense, generalizations of the second. As will be seen, these very general designs tend to require extremely large numbers of replicates and plots per block, and so they are not considered in any great detail below. An-

Received August 1, 1956.

other feature common to the first two kinds of design is that in both cases designs are possible which have only a fraction of the number of blocks and replicates necessary for the complete design. The third method of deriving designs, by addition, is unrelated to the other two; it is a slight generalization of a method given by Bose, *et al.* [3]. Apart from a few isolated examples, none of the designs given here appears in the tables of Bose, *et al.* [1].

3. Complete and incomplete groups. The first method, that of complete and incomplete groups, arises in the following manner. Consider a design with m groups of n members each, where $mn = v$. Let each block of the design contain u complete groups ($1 \leqq u \leqq m - 1$) and h members from one other group ($1 \leqq h \leqq n - 1$), the design containing sufficient blocks that the h extra treatments shall be selected in all possible ways from each of the $(m - u)$ groups not wholly represented. This necessitates ${}^nC_h(m - u)$ blocks with the same u groups, and thus ${}^mC_u{}^nC_h(m - u)$ blocks in all. The complete design thus has ${}^{m-1}C_u{}^nC_h[(nu + h)/n]$ replicates of mn treatments on ${}^mC_u{}^nC_h(m - u)$ blocks with $(nu + h)$ plots per block.

By the method of its construction the design is group divisible, save where $u = m - 1$ and $h = n - 1$ simultaneously, in which case it degenerates into a totally balanced incomplete block design. Ignoring this case, it is seen that the other parameters of the first kind are $\lambda_1 = {}^{m-1}C_u\{u{}^nC_h + {}^{n-2}C_{h-2}\}$ and $\lambda_2 = {}^{m-2}C_{u-1}\{(u - 1){}^nC_h + 2^{n-1}C_{h-1}\}$. Thus the design satisfies the conditions $r > \lambda_1$, $rk > \lambda_2 v$, where r is the number of replicates and k the number of plots per block, and so is regular.

As an illustration of the construction of the design consider the case $m = 2$, $u = 1$, $n = 4$, $h = 2$, giving a design with 9 replicates of 8 treatments on 12 blocks with 6 plots in each, with $\lambda_1 = 7$, $\lambda_2 = 6$. The design is as follows. The two groups are ABCD and EFGH and the blocks are given by columns.

A	A	A	A	A	A	E	E	E	E	E	E
B	B	B	B	B	B	F	F	F	F	F	F
C	C	C	C	C	C	G	G	G	G	G	G
D	D	D	D	D	D	H	H	H	H	H	H
E	E	E	F	F	G	A	A	A	B	B	C
F	G	H	G	H	H	B	C	D	C	D	D

All designs of this type with $r \leqq 10$ are shown in Table I.

In certain circumstances it is possible to obtain a regular design with the same m, u, n and h without taking as many blocks as would be implied by the designs just described. For example, when $m = 3$, $u = 1$, designs with just half the blocks are possible by taking blocks with the first group of treatments complete and the second incomplete, the second complete and the third incomplete, and the third complete and the first incomplete. There are then ${}^nC_h[(n + h)/n]$ replicates of $3n$ treatments on $3{}^nC_h$ blocks with $(n + h)$ plots per block, $\lambda_1 = {}^nC_h + {}^{n-2}C_{h-2}$, $\lambda_2 = {}^{n-1}C_{h-1}$. As an illustration, consider $n = 4$, $h = 1$. The design is as follows, and here $r = k = 5$, $v = b = 12$, $\lambda_1 = 4$,

$\lambda_2 = 1$. The groups are ABCD, EFGH, and IJKL, and the blocks are given by columns.

A	A	A	A	E	E	E	E	I	I	I	I
B	B	B	B	F	F	F	F	J	J	J	J
C	C	C	C	G	G	G	G	K	K	K	K
D	D	D	D	H	H	H	H	L	L	L	L
E	F	G	H	I	J	K	L	A	B	C	D

All designs of this type with $r \leqq 10$ are given in Table II.

Half designs are possible on this same principle for any odd value of m when $u = 1$; in such designs those blocks are taken in which one group is complete and another incomplete, but not conversely for the same pair of groups. As an illustration, $m = 5$, $n = 3$, $h = 1$ gives the following design with 8 replicates of 15 treatments on 30 blocks with 4 plots each, $\lambda_1 = 6$, $\lambda_2 = 1$:

A	A	A	D	D	D	G	G	G	J	J	J	M	M	M	A	A	A	G	G	G	M	M	M	D	D	D	J	J	J
B	B	B	E	E	E	H	H	H	K	K	K	N	N	N	B	B	B	H	H	H	N	N	N	E	E	E	K	K	K
C	C	C	F	F	F	I	I	I	L	L	L	O	O	O	C	C	C	I	I	I	O	O	O	F	F	F	L	L	L
D	E	F	G	H	I	J	K	L	M	N	O	A	B	C	G	H	I	M	N	O	D	E	F	J	K	L	A	B	C

The groups are ABC, DEF, GHI, JKL, and MNO, and the blocks are given by columns. For $u = 1$ all the possible half designs with 10 or fewer replicates are given below:

m	n	h	r	v	b	k	λ_1	λ_2	E_1	E_2
5	2	1	6	10	20	3	4	1	0.89	0.68
5	3	1	8	15	30	4	6	1	0.94	0.70
5	4	1	10	20	40	5	8	1	0.96	0.71
5	3	2	10	15	30	5	8	2	0.96	0.80
7	2	1	9	14	42	3	6	1	0.89	0.65

Designs of this nature with $u \neq 1$ are also sometimes possible, but there is only one with 10 or fewer replicates. This is given by $m = 5$, $u = 2$, $n = 2$, $h = 1$, and gives rise to the following design with 10 replicates of 10 treatments on 20 blocks with 5 plots each, $\lambda_1 = 8$, $\lambda_2 = 4$, $E_1 = 0.96$, $E_2 = 0.87$:

A	A	C	C	E	E	G	G	I	I	A	A	C	C	E	E	G	G	I	I
B	B	D	D	F	F	H	H	J	J	B	B	D	D	F	F	H	H	J	J
E	E	G	G	I	I	A	A	C	C	C	C	E	E	G	G	I	I	A	A
F	F	H	H	J	J	B	B	D	D	D	D	F	F	H	H	J	J	B	B
C	D	E	F	G	H	I	J	A	B	G	H	I	J	A	B	C	D	E	F

The groups are AB, CD, EF, GH, and IJ, and the blocks are given by columns.

A valid design is obtained by taking half the blocks, the first ten as the design is written down. The design has 5 replicates of 10 treatments on 10 blocks of 5 plots each, $\lambda_1 = 4$, $\lambda_2 = 2$, $E_1 = 0.96$, $E_2 = 0.87$.

Half designs are also possible for other cases where $n = 2$, $h = 1$, $m = 2u +$

TABLE I

Regular group divisible designs formed by the method of complete and incomplete groups

m	u	n	h	r	v	b	k	λ_1	λ_2	E_1	E_2
2	1	3	1	4	6	6	4	3	2	0.94	0.87
2	1	4	1	5	8	8	5	4	2	0.96	0.85
2	1	4	2	9	8	12	6	7	6	0.96	0.94
2	1	5	1	6	10	10	6	5	2	0.97	0.85
2	1	6	1	7	12	12	7	6	2	0.98	0.84
2	1	7	1	8	14	14	8	7	2	0.98	0.84
2	1	8	1	9	16	16	9	8	2	0.99	0.83
2	1	9	1	10	18	18	10	9	2	0.99	0.83
3	1	2	1	6	6	12	3	4	2	0.89	0.76
3	1	3	1	8	9	18	4	6	2	0.94	0.77
3	1	3	2	10	9	18	5	8	4	0.96	0.86
3	1	4	1	10	12	24	5	8	2	0.98	0.78
3	2	3	1	7	9	9	7	6	5	0.98	0.96
3	2	4	1	9	12	12	9	8	6	0.99	0.96
4	1	2	1	9	8	24	3	6	2	0.89	0.71

TABLE II

Regular group divisible designs with half the blocks for complete and incomplete groups with $m = 3$, $u = 1$

n	h	r	v	b	k	λ_1	λ_2	E_1	E_2
2	1	3	6	6	3	2	1	0.89	0.76
3	1	4	9	9	4	3	1	0.94	0.77
3	2	5	9	9	5	4	2	0.96	0.86
4	1	5	12	12	5	4	1	0.96	0.77
4	2	9	12	18	6	7	3	0.96	0.87
4	3	7	12	12	7	6	3	0.98	0.90
5	1	6	15	15	6	5	1	0.97	0.77
5	4	9	15	15	9	8	4	0.99	0.93
6	1	7	18	18	7	6	1	0.98	0.77
7	1	8	21	21	8	7	1	0.98	0.77
8	1	9	24	24	9	8	1	0.99	0.76
9	1	10	27	27	10	9	1	0.99	0.76

1. Putting $u = 3$ and 4 respectively gives designs with 7 replicates of 14 treatments on 14 blocks of 7 plots each, $\lambda_1 = 6$, $\lambda_2 = 3$, $E_1 = 0.98$, $E_2 = 0.91$ and 9 replicates of 18 treatments on 18 blocks of 9 plots each, $\lambda_1 = 8$, $\lambda_2 = 4$, $E_1 = 0.99$, $E_2 = 0.94$.

Further, if there is a balanced incomplete block design of n treatments with h plots per block which is not unreduced, or if the unreduced design with these parameters is resolvable, a regular design may be possible without taking all possible blocks. The only design of this type of practicable size appears to be

that with $m = 2$, $u = 1$, $n = 7$, $h = 3$, which gives rise to a design with 10 replicates of 14 treatments on 14 blocks with 10 plots in each as follows:

A − G with HIJ, HKL, HMN, IKM, ILN, JKN, JLM;
H − N with ABC, ADE, AFG, BDF, BEG, CDG, CEF.

$\lambda_1 = 8$, $\lambda_2 = 6$, $E_1 = 0.98$, $E_2 = 0.96$

Other designs, of a similar type but not necessarily containing any complete groups, can be obtained as follows. Let each block contain h_j members from the jth group of treatments in all possible ways, there being m_j groups of treatments. If j goes from 1 to s we have $k = \sum_{j=1}^{s} h_j m_j$, and the blocks can be divided into sets such that each set contains $\prod_{j=1}^{s} {}^{n}C_{h_j}$ blocks, where h_j occurs m_j times. The number of such sets of blocks is thus $m!/\prod_{j=1}^{s} m_j\,!$, where

$$\sum_{j=1}^{s} m_j = m.$$

The design contains r replicates of mn treatments on $(m!\prod {}^{n}C_{h_j})/\prod m_j$ blocks with $\sum h_j m_j$ plots per block, where summations and products run from 1 to s and the term $\prod {}^{n}C_{h_j}$ contains $h_j\, m_j$ times. Further,

$$\lambda_1 = \sum_{j=1}^{s} {}^{n-2}C_{h_j-2} \left(\prod_{j' \neq j} {}^{n}C_{h_{j'}} \right)$$

and

$$\lambda_2 = \sum_{j \neq j'} {}^{n-1}C_{h_j-1}\, {}^{n-1}C_{h_{j'}-1} \left(\prod_{j'' \neq j, j'} {}^{n}C_{h_{j''}} \right),$$

where h_j occurs m_j times.

Thus, if h_j is a constant, the design is semi-regular; otherwise it is regular. The only case with $r \leqq 10$, $k > 2$ is given by $m = 2$, $n = 3$, $s = 2$, $m_1 = m_2 = 1$, $h_1 = 2$, $h_2 = 1$, giving 9 replicates of 6 treatments on 18 blocks with 3 plots per block, $\lambda_1 = 3$, $\lambda_2 = 4$, $E_1 = 0.78$, $E_2 = 0.81$. The design is:

A	A	A	A	A	A	B	B	B	D	D	D	D	D	D	E	E	E
B	B	B	C	C	C	C	C	C	E	E	E	F	F	F	F	F	F
D	E	F	D	E	F	D	E	F	A	B	C	A	B	C	A	B	C

The groups are ABC and DEF, and the blocks are given by columns.

As before, only a fraction of the total number of blocks may be needed and two designs derived in this fashion have respectively $m = 3$, $n = 3$, $s = 3$, $m_1 = m_2 = m_3 = 1$, $h_1 = 2$, $h_2 = 1$, $h_3 = 0$, giving a half design with 9 replicates of 9 treatments on 27 blocks and 3 plots per block, $\lambda_1 = 3$, $\lambda_2 = 2$, and $m = 4$, $n = 4$, $s = 3$, $m_1 = m_2 = 1$, $m_3 = 2$, $h_1 = 2$, $h_2 = 1$, $h_3 = 0$, giving a 1/6 design with 9 replicates of 16 treatments on 48 blocks with 3 plots per block, $\lambda_1 = 2$, $\lambda_2 = 1$. Designs with these last two sets of parameters are given by Bose, *et al.* [1], but the blocks comprising the designs are different in each case, even though the efficiency factors are unaltered.

4. Designs with groups arranged in sets. In certain designs the groups of treatments may be arranged in sets in such a fashion that treatments from

groups within a set concur in one way while treatments from groups in different sets concur in another way. At first sight this would appear to lead to designs with three associate-classes, and in general it does, but in many particular cases the treatments from different groups concur the same number of times whether or not the groups are in the same set In such a case the property of group divisibility ensures that there are only two associate-classes. The simplest of these designs, those with $2n$ plots per block, are derived below.

Consider a design with $2n^2$ treatments in $2n$ groups of n members each, the design having $2n$ plots per block. Divide the groups of treatments into two sets, set 1 containing the first n groups and set 2 the remainder. The design then has blocks of the following kinds:

(i) Blocks containing two complete groups from set 1 or set 2, there being $[n(n-1)]/2$ from each set and thus $n(n-1)$ in all. There are thus $(n-1)$ replicates of each treatment in these blocks.

(ii) Blocks containing one complete group from set 1 and one member from each group in set 2, or conversely. These blocks are sufficient in number that each group from one set occurs once with every treatment from the other and that the groups of a set occur equally frequently. This necessitates n^2 blocks with complete groups from one set and one member from each group of the other, and thus $2n^2$ blocks in all. Further, in order that each treatment shall occur once and once only with all treatments in the same set but different groups, n must be such that there are $(n-1)$ orthogonal Latin squares of side n. Each treatment is replicated $2n$ times in these blocks.

The complete design thus has $(3n-1)$ replicates and $n(3n-1)$ blocks.

In blocks of the first kind, each treatment concurs $(n-1)$ times with treatments of its own group, once with treatments of the other groups of its own set, and not at all with treatments of the other set. In blocks of the second kind, each treatment concurs n times with treatments of its own group, once with treatments of the other groups in its own set, and twice with treatments of the other set. Thus the design is group divisible with $\lambda_1 = 2n-1$, $\lambda_2 = 2$, and so, further, is regular.

$n = 2$ and $n = 3$ give the only examples with 10 or fewer replicates, these having respectively 5 replicates of 8 treatments on 10 blocks with 4 plots each, $\lambda_1 = 3$, $E_1 = 0.90$, $E_2 = 0.85$, and 8 replicates of 18 treatments on 24 blocks with 6 plots each, $\lambda_1 = 5$, $E_1 = 0.94$, $E_2 = 0.87$. The design for $n = 3$ is:

Blocks of the first kind						Blocks of the second kind											
A	A	D	J	J	M	A	A	A	D	D	D	G	G	G	J	J	J	M	M	M	P	P	P
B	B	E	K	K	N	B	B	B	E	E	E	H	H	H	K	K	K	N	N	N	Q	Q	Q
C	C	F	L	L	O	C	C	C	F	F	F	I	I	I	L	L	L	O	O	O	R	R	R
D	G	G	M	P	P	J	K	L	J	K	L	J	K	L	A	B	C	A	B	C	A	B	C
E	H	H	N	Q	Q	M	N	O	N	O	M	O	M	N	D	E	F	E	F	D	F	D	E
F	I	I	O	R	R	P	Q	R	R	P	Q	Q	R	P	G	H	I	I	G	H	H	I	G

The groups are ABC, DEF, GHI, JKL, MNO, and PQR, and the blocks are given by columns.

If there are $3n$ groups of n members each instead of $2n$ groups, a design with $2n$ plots per block is possible in a similar fashion. Here, however, only the second kind of block described above is used, and so the design is, in a sense, a complete and incomplete group design as understood in the last section. The three possible pairs of sets of groups all have to be considered, thus giving rise to a design with $6n^2$ blocks and $4n$ replicates. If two treatments belong to different groups, whether of the same set or not, they concur twice, i.e. $\lambda_2 = 2$; further, $\lambda_1 = 2n$, and the design is thus regular group divisible. The design with $n = 2$ is the only one with 10 replicates or fewer, and has in fact 8 replicates of 12 treatments on 24 blocks with 4 plots each, $\lambda_1 = 4$, $\lambda_2 = 2$, $E_1 = 0.88$, $E_2 = 0.81$. The design is:

A A C C E E G G A A C C I I K K E E G G I I K K
B B D D F F H H B B D D J J L L F F H H J J L L
E F E F A B A B I J I J A B A B I J I J E F E F
G H H G C D D C K L L K C D D C K L L K G H H G

The groups are AB, CD, EF, GH, IJ, and KL, and the blocks are given by columns.

In the same way that, for complete and incomplete groups half designs are possible with $m = 3$, $u = 1$, so also are half designs possible here by means of the same device, i.e., with complete groups from the first set and incomplete from the second, and so on in a cyclic fashion only. $n = 2$ gives a design with 4 replicates of 12 treatments on 12 blocks with 4 plots each, and this design is given by Bose, *et al.* [1]. In general the design has $2n$ replicates of $3n^2$ treatments on $3n^2$ blocks with $2n$ plots each, $\lambda_1 = n$, $\lambda_2 = 1$. For $n = 3$, 4, or 5 respectively, the design thus has 6, 8, or 10 replicates and plots per block, and it has 27, 48, or 75 treatments and blocks; $\lambda_1 = 3$, 4, or 5; $E_1 = 0.92$, 0.94, or 0.95; $E_2 = 0.85$, 0.88, or 0.90.

Designs of this kind are possible with more than $2n$ plots per block, but the numbers of replicates and plots per block very soon give designs which are beyond the bounds of practicality. Even the smallest designs with $3n^2$ treatments and $3n$ plots per block have 12 replicates. However, the smallest design with $2n^2$ treatments and $3n$ plots per block, that with $n = 3$, gives a more practical design. This design, which has 9 replicates of 18 treatments on 18 blocks with 9 plots each, $\lambda_1 = 6$, $\lambda_2 = 3$, $E_1 = 0.96$, $E_2 = 0.91$, arises in each of the following forms:

A	A	A	A	A	A	D	D	D	J	J	J	J	J	J	M	M	M
B	B	B	B	B	B	E	E	E	K	K	K	K	K	K	N	N	N
C	C	C	C	C	C	F	F	F	L	L	L	L	L	L	O	O	O
D	D	D	G	G	G	G	G	G	M	M	M	P	P	P	P	P	P
E	E	E	H	H	H	H	H	H	N	N	N	Q	Q	Q	Q	Q	Q
F	F	F	I	I	I	I	I	I	O	O	O	R	R	R	R	R	R
J	K	L	J	K	L	J	K	L	A	B	C	A	B	C	A	B	C
M	N	O	N	O	M	O	M	N	D	E	F	E	F	D	F	D	E
P	Q	R	R	P	Q	Q	R	P	G	H	I	I	G	H	H	I	G

or

G	G	G	D	D	D	A	A	A	P	P	P	M	M	M	J	J	J
H	H	H	E	E	E	B	B	B	Q	Q	Q	N	N	N	K	K	K
I	I	I	F	F	F	C	C	C	R	R	R	O	O	O	L	L	L
K	J	J	K	J	J	K	J	J	B	A	A	B	A	A	B	A	A
L	L	K	L	L	K	L	L	K	C	C	B	C	C	B	C	C	B
N	M	M	M	M	N	M	N	M	E	D	D	D	D	E	D	E	D
O	O	N	O	N	O	N	O	O	F	F	E	F	E	F	E	F	F
Q	P	P	P	Q	P	P	P	Q	H	G	G	G	H	G	G	G	H
R	R	Q	Q	R	R	R	Q	R	I	I	H	H	I	I	I	H	I

The groups are ABC, DEF, GHI, JKL, MNO, and PQR, and the blocks are given by columns.

These two designs illustrate the principles on which the general designs of this type are derived, *viz.*, either two complete groups from one set and one member from each group of the other set or one complete group from one set and two members from each group of the other set.

5. Designs derivable by addition. Bose, *et al.* [3] describe designs derivable by addition of further blocks to a balanced incomplete block design with

TABLE III

Regular group divisible designs derivable by addition

v	k	m	n	r_1	b_1	Type	Rep.	r_2	b_2	Type	Rep.	r	b	λ_1	λ_2	E_1	E_2
6	3	2	3	5	10	T	1	1	2	S	1	6	12	3	2	0.83	0.77
6	3	2	3	5	10	T	1	1	2	S	2	7	14	4	2	0.86	0.73
6	3	2	3	5	10	T	1	1	2	S	3	8	16	5	2	0.88	0.70
6	3	2	3	5	10	T	1	1	2	S	4	9	18	6	2	0.89	0.67
6	3	3	2	5	10	T	1	2	4	SR	1	7	14	2	3	0.76	0.81
6	3	3	2	5	10	T	1	2	4	SR	2	9	18	2	4	0.74	0.81
6	3	3	2	6	12	R	1	2	4	SR	1	8	16	4	3	0.83	0.79
8	4	2	4	7	14	T	1	1	2	S	1	8	16	4	3	0.88	0.84
8	4	2	4	7	14	T	1	1	2	S	2	9	18	5	3	0.89	0.82
8	4	2	4	7	14	T	1	1	2	S	3	10	20	6	3	0.90	0.80
8	4	4	2	7	14	T	1	3	6	S	1	10	20	6	4	0.90	0.85
8	4	4	2	5	10	R	1	3	6	S	1	8	16	6	3	0.94	0.83
8	4	4	2	6	12	SR	1	3	6	S	1	9	18	3	4	0.83	0.86
9	6	3	3	8	12	T	1	2	3	S	1	10	15	7	6	0.95	0.93
10	4	2	5	6	15	T	1	4	10	R	1	10	25	5	2	0.88	0.76
10	4	5	2	6	15	T	1	4	10	S	1	10	25	6	3	0.90	0.82
10	5	5	2	9	18	T	1	1	2	S	1	10	20	5	4	0.90	0.88
12	3	3	4	4	16	SR	1	3	12	R	2	10	40	4	1	0.80	0.64

r_1, b_1 and r_2, b_2 are the numbers of replicates and blocks in the two designs which are added to make the final design given here, the number of complete replications of these initial designs being given in the appropriate column. The type of initial design is also given, T for totally balanced incomplete block designs and S, SR, and R respectively for singular, semi-regular, and regular group divisible incomplete block designs.

$v = mk$, m blocks giving a complete replicate, and designs derivable by taking together the blocks of two group divisible designs with the same v and k. However, it is possible to add blocks to a balanced design even with $v \neq mk$, or $v = mk$ when m of the balanced design blocks do not give a complete replicate, or to add a group divisible design to a balanced design, all of which amount to the same thing. New designs ($r \leqq 10$) derived by these and other addition methods, are given in Table III. An example of this type with $v = mk$ is given by the following design with 7 replicates of 6 treatments in 2 groups of 3 on 14 blocks with 3 plots each, $\lambda_1 = 4$, $\lambda_2 = 2$. The groups are ABC, and DEF, and the blocks are given by columns.

A	A	A	A	A	B	B	B	C	C	A	D	A	D
B	B	C	D	E	C	D	E	D	D	B	E	B	E
C	D	F	E	F	E	F	F	E	F	C	F	C	F

Totally balanced incomplete block | Partially balanced (singular, disconnected)

An example with $v \neq mk$ is the design with 10 replicates of 10 treatments in 2 groups of 5 on 25 blocks with 4 plots each, $\lambda_1 = 5$, $\lambda_2 = 2$. The design is:

A	B	A	A	A	B	C	D	D	A	B	C	A	B	C	A	A	A	A	B	F	F	F	F	G
B	C	C	B	D	E	D	E	F	E	F	F	F	D	E	B	B	B	C	C	G	G	G	H	H
C	E	G	I	E	G	I	H	G	G	H	G	H	G	H	C	C	D	D	D	H	H	I	I	I
D	F	H	J	F	I	J	I	J	J	J	I	I	H	J	D	E	E	E	E	I	J	J	J	J

Totally balanced incomplete block | Partially balanced (disconnected)

The groups are ABCDE and FGHIJ, and the columns represent the blocks. This design illustrates the point made by Bose, *et al.* [3] that it does not matter if a design is disconnected if it is added to another design to form a new one.

REFERENCES

[1] R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, "Tables of partially balanced designs with two associate classes," *North Carolina Agricultural Experimental Station Technical Bulletin*, No. 107 (1954).

[2] R. C. Bose and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[3] R. C. Bose, S. S. Shrikhande, and K. N. Bhattacharya, "On the construction of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 167–195.

ON PARTIALLY BALANCED LINKED BLOCK DESIGNS

By J. Roy and R. G. Laha

Indian Statistical Institute

1. Summary. The computations in the analysis of any equireplicate design can be carried out very easily if the number of treatments common to any two blocks is constant. A design with this property is called a Linked Block (LB) design and was introduced by Youden [9]. It is well known that for a Balanced Incomplete Block (BIB) design to have a constant number of treatments in common between any two blocks, it is necessary and sufficient that it is symmetric, that is, the number of blocks is equal to the number of treatments. In this paper, necessary and sufficient conditions are derived for any design with a given treatment-structure matrix to be of the LB type and the results applied to Partially Balanced Incomplete Block (PBIB) designs. Finally a list is prepared of all LB designs in the class of two-associate PBIB designs enumerated by Bose, Shrikhande and Clatworthy [2].

2. Introduction. An arrangement of v treatments in b blocks, each of k plots, $k < v$, such that each treatment occurs at most once in any block and altogether in r blocks is called an incomplete block design and denoted by $D(v, b, k, r)$. Obviously $vr = bk$. A design with $b = v$ is called symmetric. A $D(v, b, k, r)$ is completely characterised by its 'incidence matrix' $N = ((n_{ij}))$ where $n_{ij} = 1$ if the ith treatment occurs in the jth block and $n_{ij} = 0$ otherwise $i = 1, 2, \cdots, v, j = 1, 2, \cdots, b$. The matrix $\Lambda = NN'$ where $\lambda_{ii} = r$ and $\lambda_{ij} =$ the number of blocks in which the treatments i and j occur together $i \neq j = 1, 2, \cdots, v$ is called the 'treatment-structure matrix' of the design. A design is balanced if $\lambda_{ij} = \lambda$ for all $i \neq j$. The design obtained from $D(v, b, k, r)$ by considering its blocks as treatments and treatments as blocks is called its dual. The number of treatments common to the ith and the jth blocks will be denoted by $\mu_{ij}, i, j = 1, 2, \cdots, b$. The matrix $M = N'N$ has been called the 'structural matrix' by Connor [5] and Connor and Hall [7]. We shall however call M the 'block-structure matrix'. A design is of the LB type if $\mu_{ij} = \mu$ for all $i \neq j$. Obviously a design is of the LB type if and only if its dual is balanced. For a definition of PBIB design the reader is referred to Bose and Shimamoto [4].

3. Conditions for a design with given treatment-structure matrix to be of the LB type. We first give two lemmas which are useful in deriving necessary and sufficient requirements on the treatment-structure matrix.

Lemma 3.1. *If A and B are two matrices of the form $m \times n$ and $n \times m$ respectively, the non-zero latent roots of AB are identical with those of BA and if corresponding to a latent root θ, ξ is a latent vector of AB, $\eta = \xi A$ will be a latent vector*

Received December 12, 1955; revised October 10, 1956.

of BA corresponding to the same root θ. If θ is a non-zero repeated latent root of AB of multiplicity r, it is so for BA also.

PROOF. If θ is a non-zero latent root of AB of multiplicity r, we can always find r linearly independent vectors ξ_i satisfying $\xi_i AB = \theta\xi_i$ $(i = 1, 2, \cdots, r)$ Post-multiplication by A gives $\eta_i BA = \theta\eta_i$ where $\eta_i = \xi_i A$. That η_i's are linearly independent follows from the linear independence of ξ_i's because for any set of constants c_i's, $\sum c_i\eta_i B = \theta\sum c_i\xi_i$. This also shows that θ is also a latent root of BA of multiplicity r and η_i's are a corresponding set of linearly independent latent vectors.

LEMMA 3.2. *The necessary and sufficient condition for a symmetric matrix A of order n to have all its diagonal elements equal and all its off-diagonal elements equal is that it has only two latent roots, one of multiplicity $(n-1)$ and the vector $(1, 1, \cdots, 1)$ is a latent vector corresponding to the other latent root.*

PROOF. Necessity is obvious. To prove that the conditions are sufficient, let us write

$$\alpha = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}\right).$$

Since A is symmetric there exists an orthogonal matrix

$$C = \begin{bmatrix} \alpha \\ P \end{bmatrix}$$

such that

$$CAC' = \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 I_{n-1} \end{bmatrix},$$

where θ_1 and θ_2 are the latent roots, θ_2 with multiplicity $(n-1)$ and I_n is the identity matrix of order n. Premultiplying by C' and post-multiplying by C we get

$$A = \theta_1\alpha'\alpha + \theta_2 P'P = (\theta_1 - \theta_2)\alpha'\alpha + \theta_2 I_n,$$

which has diagonal elements equal to $\{\theta_1 + (n-1)\theta_2\}/n$ and off-diagonal elements equal to $(\theta_1 - \theta_2)/n$.

We are now in a position to prove

THEOREM 3.1. *The necessary and sufficient condition for a design $D(v, b, k, r)$ to be of the* LB *type is that $k - \mu$ is a latent root of the treatment-structure matrix Λ, of multiplicity $(b-1)$ where $\mu = k(r-1)/(b-1)$.*

PROOF. The necessity is obvious. To prove the sufficiency of the conditions, let us write N for the incidence matrix of the given design. Then we have to show that the block-structure matrix $M = N'N$ has all off-diagonal elements equal. Since it is given that $k - \mu$ is a latent root of multiplicity $(b-1)$ of the treatment-structure matrix $\Lambda = NN'$ by Lemma 3.1 it will be so for $M = N'N$ also. Again since the total of each column of Λ is rk, $\epsilon = (1, 1, \cdots, 1)$ is a latent vector of Λ corresponding to the latent root rk. Again by Lemma 3.1, $\epsilon N = (k, k,$

$\cdots, k)$ is a latent vector of M corresponding to the latent root rk, therefore so is also the vector $(1, 1, \cdots, 1)$. Thus M satisfies all the conditions of Lemma 3.2. Hence it has all diagonal elements equal to k and all off-diagonal elements equal to μ. But since $\mu = \sum_{t=1}^{v} n_{ti}n_{tj}$ for all $i \neq j$ and $n_{ti} = 1$ or 0, μ must be integral. The number of treatments common to any two blocks is thus μ.

COROLLARY 3.1. *If the treatment-structure matrix of any design $D(v, b, k, r)$ has only two non zero latent roots, rk and $k(b-r)/(b-1)$ and rk is not a repeated root, then the design must be of the* LB *type.*

PROOF. If t is the multiplicity of the root $k(b-r)/(b-1)$ equating the sum of the diagonal elements of the treatment-structure matrix to the sum of the latent roots, we get $t = (b-1)$. Hence the result.

COROLLARY 3.2. *If $D(v, b, k, r)$ is balanced, the necessary and sufficient condition that it is of the* LB *type is that $v = b$.*

PROOF: Since the design is balanced, its treatment-structure matrix has a latent root of multiplicity $v-1$, while if it is of the LB type the multiplicity must be $b-1$. Hence $b = v$.

4. Partially balanced linked block designs. We now apply the results of Section 3 to the special case of PBIB design $D(v, b, k, r)$ having m associate classes with parameters n_i, λ_i, $p_{js}^{i}(i, j, s = 1, 2, \cdots, m)$ as defined in Bose and Shimamoto [4]. It follows from the results of Connor and Clatworthy [6] that latent roots other than rk of the treatment-structure matrix for such a design are, except for repetitions, the same as the latent roots of the reduced matrix $A = ((a_{ij}))$ of order m where

$$(4.1)\qquad \begin{cases} a_{ii} = r + \lambda_1 p_{i1}^{i} + \lambda_2 p_{i2}^{i} + \cdots + \lambda_m p_{im}^{i} - \lambda_i n_i, \\ a_{ij} = \quad \lambda_1 p_{i1}^{j} + \lambda_2 p_{i2}^{j} + \cdots + \lambda_m p_{im}^{j} - \lambda_i n_i, \\ \qquad\qquad\qquad\qquad\qquad\qquad i \neq j = 1, 2, \cdots m. \end{cases}$$

Hence we have the

THEOREM 4.1. *The necessary and sufficient condition for a* PBIB *design with m associate classes to be of the* LB *type is that the matrix A defined in* (4.1) *has only one non zero latent root $k(b-r)/(b-1)$.*

COROLLARY 4.1. *The necessary and sufficient condition for a* PBIB *design with two associate classes to be of the* LB *type is that*

$$(4.2)\qquad a_{11}a_{22} - a_{21}a_{12} = 0,$$

$$(4.3)\qquad a_{11} + a_{22} = k(b-r)/(b-1),$$

where

$$(4.4)\qquad \begin{cases} a_{11} = r + \lambda_1 p_{11}^{1} + \lambda_2 p_{12}^{1} - \lambda_1 n_1, \\ a_{12} = \quad \lambda_1 p_{11}^{2} + \lambda_2 p_{12}^{2} - \lambda_1 n_1, \\ a_{21} = \quad \lambda_1 p_{21}^{1} + \lambda_2 p_{22}^{1} - \lambda_2 n_2, \\ a_{22} = r + \lambda_1 p_{21}^{2} + \lambda_2 p_{22}^{2} - \lambda_2 n_2. \end{cases}$$

PROOF. The matrix $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ cannot have two equal latent roots.

We shall now apply this result to some special types of PBIB designs with two associate classes.

4.1 *Group Divisible* (GD) *Designs.* A GD design as defined by Bose and Connor [1] is specified by the parameters

$$v = mn, \qquad n_1 = (n-1), \qquad n_2 = m(n-1),$$

$$((p_{ju}^1)) = \begin{bmatrix} n-2 & 0 \\ 0 & n(m-1) \end{bmatrix}, \qquad ((p_{ju}^2)) = \begin{bmatrix} 0 & n-1 \\ n-1 & n(m-2) \end{bmatrix}.$$

They have classified the GD designs as (i) Singular when $r = \lambda_1$ (ii) Semi-Regular when $r > \lambda_1$ and $rk = v\lambda_2$ and (iii) Regular when $r > \lambda_1$ and $rk > v\lambda_2$. They have also shown that for a Singular GD design, $b \geqq m$ and for a Semi-Regular GD design $b \geqq v - m + 1$. In order that a GD design may be of the LB type the condition (4.2) gives

$$(r - \lambda_1)(rk - v\lambda_2) = 0,$$

so that a Regular GD design is never of the LB type. The condition (4.3) gives for the case $r = \lambda$, $b = m$ and for the case $rk = v\lambda_2$, $b = v - m + 1$. We now summarise these results in the form of the

THEOREM 4.2. *A Regular* GD *design cannot be of the* LB *type. The necessary and sufficient condition for a Singular* GD *to be of the* LB *type is that* $b = m$ *and that for a Semi-Regular* GD *is that* $b = v - m + 1$.

Bose and Connor [1] have shown that a Singular GD design can always be derived from a BIB design with m treatments by replacing each treatment by a group of n treatments. Hence the condition for a Singular GD design to be of the LB type is that the BIB design from which it is generated should be symmetric.

4.2 *Triangular Designs.* In a Triangular design (Bose and Shimamoto [4])

$$v = \tfrac{1}{2}n(n-1), \qquad n_1 = 2(n-2), \qquad n_2 = \tfrac{1}{2}(n-2)(n-3),$$

$$((p_{ju}^1)) = \begin{bmatrix} n-2 & n-3 \\ n-3 & \tfrac{1}{2}(n-3)(n-4) \end{bmatrix}, \quad ((p_{ju}^2)) = \begin{bmatrix} 4 & 2(n-4) \\ 2(n-4) & \tfrac{1}{2}(n-4)(n-5) \end{bmatrix}.$$

In order that a Triangular design may be of the LB type the condition (4.2) gives

$$(r - 2\lambda_1 + \lambda_2)\{r + (n-4)\lambda_1 - (n-3)\lambda_2\} = 0.$$

From the other condition (4.3) we get $b = n$ if $r - 2\lambda_1 + \lambda_2 = 0$ and $b = \tfrac{1}{2}(n-1)(n-2)$ if $r + (n-4)\lambda_1 - (n-3)\lambda_2 = 0$. We thus get the

THEOREM 4.3. *The necessary and sufficient condition for a Triangular design to be of the* LB *type is that either* (i) $r = 2\lambda_1 - \lambda_2$ *and* $b = n$ *or* (ii) $r = (n-3)\lambda_2 - (n-4)\lambda_1$ *and* $b = \tfrac{1}{2}(n-1)(n-2)$.

It is interesting to note that in case (i) if $r = 2$ we get the Triangular Singly Linked Block (TSLB) designs and in case (ii) if $r = n - 2$ the Triangular Doubly Linked Block (TDLB) designs as defined by Bose and Shimamoto

[4]. The results of Theorem 4.3 may thus be considered as generating the class of Triangular Multiply Linked Block designs.

4.3 *Two associate* PBIB *designs with* $k > r \geqq 2$ *and* $\lambda_1 = 1$, $\lambda_2 = 0$ (*Simple* PBIB). Bose and Clatworthy [3] have shown that all designs of this class are characterized by the parameters

$$v = k[(r-1)(k-1)+t]/t, \qquad b = r[(r-1)(k-1)+t]/t,$$

$$n_1 = r(k-1), \qquad n_2 = (r-1)(k-1)(k-t)/t$$

$$((p_{ij}^1)) = \begin{bmatrix} (t-1)(k-1)+k-2 & (r-1)(k-t) \\ (r-1)(k-t) & (r-1)(k-t)(k-t-1)/t \end{bmatrix}$$

$$((p_{ij}^2)) = \begin{bmatrix} rt & r(k-t-1) \\ r(k-t-1) & [(r-1)(k-1)(k-2t)+t(rt-k)]/t \end{bmatrix}$$

where $1 \leqq t \leqq r$. It is interesting to note that in this case the condition for an LB design is $t = r$ and then $\mu = 1$. Hence we have the

THEOREM 4.4. *The only* LB *designs in the class of two-associate* PBIB *designs with* $k > r \geqq 2$ *and* $\lambda_1 = 1$, $\lambda_2 = 0$ *are those which are duals of* BIB *designs in which any two treatments occur together in just one block.*

Shrikhande [8] showed that the dual of any BIB design with $\lambda = 1$ is a two associated PBIB design with $\lambda_1 = 1$, and $\lambda_2 = 0$. Our result shows that no two-associate PBIB design with $\lambda_1 = 1$ and $\lambda_2 = 0$ and $k > r \geqq 2$ can be obtained by dualising BIB designs other than those with $\lambda = 1$.

4.4 *List of two-associate* PBIB *designs of the* LB *type.* We give below a list of LB designs in the class of two associate PBIB designs enumerated by Bose, Clatworthy and Shrikhande [2]. The reference number for a design is the one used by the above authors and μ denotes other number of treatments common to any two blocks.

TABLE 4.1

List of two-associate PBIB *designs of the* LB *type.* (S = *Singular* GD, SR = *Semi-Regular* GD, Sl = *Simple*, T = *Triangular.*)

Reference No.	μ	Reference No.	μ	Reference No.	μ
S 1	2	SR 1	1	Sl 25	1
S 7	4	SR 20	1	T 1	1
S 12	3	SR 26	5	T 9	2
S 18	6	SR 32	3	T 15	3
S 22	4	SR 51	1	T 20	1
S 24	6	SR 70	1	T 22	2
S 28	8	SR 85	1	T 25	6
S 40	2	SR 89	1	T 27	5
S 41	4	Sl 4	1	T 31	1
S 46	5	Sl 9	1	T 32	1
S 77	3	Sl 17	1	T 33	1
S 81	4	Sl 18	1	T 34	2
S 89	2	Sl 21	1	T 35	1
S 111	2	Sl 22	1	T 36	1

REFERENCES

[1] Bose, R. C., and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–382.

[2] Bose, R. C., W. H. Clatworthy, and S. S. Shrikhande, *Tables of Partially Balanced Designs with Two-associate Classes*, Technical Bulletin No. 107, North Carolina Agricultural Experiment Station (1954).

[3] Bose, R. C., and W. H. Clatworthy, "Some classes of partially balanced designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 212–232.

[4] Bose, R. C., and T. Shimamoto, "Classification and analysis of partially balanced incomplete block designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–184.

[5] Connor, W. S., "On the structure of balanced incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1953), pp. 57–71.

[6] Connor, W. S., and W. H. Clatworthy, "Some theorems for partially balanced designs," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 100–112.

[7] Connor, W. S., and M. Hall, "An embedding theorem for balanced incomplete block designs," *Canadian J. Math.*, Vol. 6 (1953), pp. 35–41.

[8] Shrikhande, S. S., "On the dual of some balanced incomplete block designs," *Biometrics*, Vol. 8 (1952), pp. 66–72.

[9] Youden, W. G., "Linked blocks: a new class of incomplete block designs" (Abstract), *Biometrics*, Vol. 7 (1951), p. 124.

NOTES

ON BOREL FIELDS OVER FINITE SETS

By G. Szekeres and F. E. Binet

University of Adelaide and University of Melbourne

1. Summary. It is shown that the number of Borel Fields over a set (S) of n elements is equal to the number of equivalence relations within S. This number is asymptotically equal to

$$(\beta + 1)^{-1/2} \exp \{n(\beta - 1 + \beta^{-1}) - 1\} \qquad \text{where} \qquad \beta \exp \beta = n.$$

2. Enumeration of Borel Fields over a finite set. Borel Fields are usually (e.g. Wald [8]) defined over a set of non-enumerably infinite elements: with quite trivial changes, the definition is applicable to finite sets, as follows:

Let $A, B, C, \cdots$ denote distinct subsets of a set S of n elements. $\mathfrak{B} = \{A, B, C, \cdots\}$ is called a Borel Field (BF) if and only if

(i) $\mathfrak{B}$ is not empty;

(ii) $A \varepsilon \mathfrak{B}$, $B \varepsilon \mathfrak{B}$ imply

$A \cap B \varepsilon \mathfrak{B}$, $A \cup B \varepsilon \mathfrak{B}$, $S - A \varepsilon \mathfrak{B}$.

It follows from the definition that a BF contains at least the empty set ($\varnothing$) and S, and is closed with respect to the formation of unions, intersections, and complements.

To enumerate the BF's, consider the subset $\mathfrak{P}$ consisting of all $P_m \varepsilon \mathfrak{B} (m = 1, 2, \cdots, r$; for some $r = 1, 2, \cdots, n)$ such that

$$P \neq \varnothing, \tag{1}$$

$$A \neq \varnothing, \quad A \neq P, \quad A \varepsilon \mathfrak{B} \quad \text{implies} \quad A \not\subset P; \tag{2}$$

in others words, no P contains an element of $\mathfrak{B}$ as a proper subset. It follows that

$$P_m \cap P_{m'} = \varnothing \qquad (\text{for } m \neq m') \tag{3}$$

and

$$\cup_m P_m = S. \tag{4}$$

If (3) were not true, the intersection, itself being an element of the BF and also a proper subset of a P, would involve a contradiction with (2); if (4) were not so, the complement of this union, being an element (other than $\varnothing$) of the BF and therefore not containing a subset of any other P, would itself be a P, namely P_{r+1}, contradicting the definition of $P = \{P_m\}$.

It is obvious that a BF defines a unique $\mathfrak{P}$; conversely a $\mathfrak{P}$ defines a unique BF as follows:

$$\mathfrak{B} = \{\varnothing; P_1, P_2, \cdots, P_r; \tbinom{r}{2} \text{ elements like } P_1 \cup P_2;$$

Received June 14, 1956.

$\binom{r}{2}$ elements like $P_1 \cup P_2 \cup P_3 ; \cdots ; S\}$.

Thus every BF consists of 2^r elements, the number of BF's with 2^r elements being the same as that of $\mathfrak{P}$'s with r elements. This latter, however, is known to be $\Delta^r 0^n/r!$, where $\Delta^r 0^n$ is the leading rth difference of nth powers of the non-negative integers.

It is obvious from the foregoing that the total number of BF's over S is the same as the total number of $\mathfrak{P}$'s, namely

$$\sum_{r=1}^{n} \Delta^r 0^n/r! = G_n, \tag{5}$$

say; it is also equal to the number of equivalence relations within S. It is well known that

$$\sum_{n=0}^{\infty} z^n G_n/n! = \exp(e^z - 1); \tag{6}$$

in conventional symbolic notation $G_{n+1} = (1 + G)^n$. Bell [2] gives this recurrence relation as well as several realizations of G_n. We give two further simple realizations:

First, (6) shows that G_n is the nth power-moment, around zero, of the Poisson distribution with unit parameter,

$$Pr(X = x) = (ex!)^{-1}, \qquad x = 0, 1, \cdots.$$

Second, (see, for example, Fisher [4]),

$$\Delta^r 0^n/r! = \Sigma\left\{n! \prod_{\nu=1}^{R} (\nu!)^{k_\nu} k_\nu!\right\}, \tag{7}$$

where summation takes place over all R, ν, k, such that

$$\sum_{\nu=1}^{R} \nu k_\nu = n, \tag{8}$$

$$\sum_{\nu=1}^{R} k_\nu = r. \tag{9}$$

The typical term is the number of ways n elements can be distributed corresponding to the partition of n, symbolically represented by

$$1^{k_1} 2^{k_2} \cdots \nu^{k_\nu} \cdots R^{k_R},$$

with ν, k_ν satisfying (8) and (9). Dropping the restriction due to (9), but keeping that due to (8), the sum becomes G_n.

3. Evaluation of G_n. For $n = 1$ to 20, Epstein [3] tabulates G_n, using (5).[1] He also gives an asymptotic evaluation of G_n, expressed in terms of the function $\Psi(x) = d/dx \log \Gamma(x)$ and the numbers α_n defined through the relation

$$\alpha_n \Psi(\alpha_n + 1) = n.$$

[1] For $n = 21$ to 51 an unpublished table has been prepared by Francis L. Miksa, 613 Spring Street, Aurora, Ill., U.S.A.

We shall give here a more direct asymptotic expression for G_n in terms of elementary functions; it is obtained by evaluating

$$I_n = \oint_C z^{-(n+1)} \exp(e^z)\, dz, \tag{10}$$

where C is a simple contour enclosing the origin of the z-plane. Clearly by (6) and by Cauchy's theorem,

$$G_n = \frac{n!}{2\pi i e} I_n. \tag{11}$$

To obtain an asymptotic expression for I_n, we specify C in (10) by $|z| = \beta$, with $\beta = \beta(n)$ defined by

$$\beta e^{\beta} = n; \tag{12}$$

then C intersects the positive real axis very nearly at a point where the derivative of the integrand vanishes, and the integral can be evaluated by the method of steepest descent. By a modification of Watson's Lemma (see Jeffreys [6]) it can be shown (details are given in the Appendix) that

$$\begin{aligned} G_n &= n! \exp(n\beta^{-1} - 1)\beta^{-1}\{2\pi n(\beta+1)\}^{-1/2} \\ &\times\{1 - (2\beta^4 + 9\beta^3 + 16\beta^2 + 6\beta + 2)(24n)^{-1}(\beta+1)^{-3} + 0(\beta^2 n^{-2})\}; \end{aligned} \tag{13}$$

or using Stirling's formula this simplifies to

$$\begin{aligned} G_n &= (\beta+1)^{-1/2} \exp\{n(\beta - 1 + \beta^{-1}) - 1\} \\ &\times\{1 - \beta^2(2\beta^2 + 7\beta + 10)(24n)^{-1}(\beta+1)^{-3} + 0(\beta^2 n^{-2})\} \end{aligned} \tag{14}$$

$$= (\beta+1)^{-1/2} \exp\{n(\beta - 1 + \beta^{-1}) - 1\}\{1 + 0(\beta n^{-1})\}. \tag{15}$$

These are the required asymptotic formulae. It should be mentioned that (15) can also be obtained from Epstein's result, with the help of Stirling's formula; but (14) would require the knowledge of Epstein's second asymptotic term which has not been determined explicitly in his paper.

The following table gives comparative values of $\log G_{51}$ as computed from the various asymptotic formulae:

$\log G_{51}$ (true value)	111.707033
from (14)	111.707084
from (15)	111.712500
from Epstein	111.706867

The true value was obtained from Miksa's value for G_{51} (l.c. footnote 1).

By a similar method as above it can be shown that for $r < n/\log n$

$$\Delta^r 0^n = r^n \exp\left\{\left(\frac{1}{2}\frac{n}{r} - r\right)e^{-n/r}\right\} \times\left\{1 + 0\left(\frac{1}{n}\right)\right\}. \tag{16}$$

This sharpens Jordan's result [7]

$$\lim_{n\to\infty} r^{-n}\Delta^r 0^n = 1,$$

and establishes a connection between (5) and the known formula (see, for example Bell, [2])

$$G_n = e^{-1}\sum_{r=1}^{\infty} r^n/r!.$$

Other asymptotic formulae for $\Delta^r 0^n$ have been obtained previously by Hsu [5] and by Arfwedson [1], the former being valid when $n - r = 0(n^{1/2})$, the latter when $r = Kn$, for any constant $K < 1$.

4. Appendix. From (10) we get (with $z = \beta e^{i\varphi}$)

$$I_n = i\int_{-\pi}^{\pi} \beta^{-n}\exp\{-ni\varphi + \exp(\beta e^{i\varphi})\}\,d\varphi$$

(A1)

$$= i\beta^{-n}\exp(e^{\beta})\left\{\int_{-\delta}^{+\delta} + \int_{\delta}^{\pi} + \int_{-\pi}^{-\delta}\exp(-ni\varphi + \exp(\beta e^{i\varphi}) - e^{\beta})\right\}d\varphi,$$

where $0 < \delta \leqq \pi$. We can choose

$$\delta = n^{-2/5}. \tag{A2}$$

Then we have, for $\delta \leqq \varphi \leqq \pi$,

$$\begin{aligned}|\exp\{-ni\varphi + \exp(\beta e^{i}\varphi) - e^{\beta}\}| &\leqq \exp(e^{\beta\cos\delta} - e^{\beta})\\ &< \exp\{-\tfrac{1}{2}\beta e^{\beta}(1-\cos\delta)\}\\ &< \exp\{-cn^{1/5}\},\end{aligned}$$

for a suitably chosen constant $c > 0$. Hence

$$\left|\int_{\delta}^{\pi}\right| < \pi\exp\{-cn^{1/5}\} \tag{A3}$$

and similarly

$$\left|\int_{-\pi}^{-\delta}\right| < \pi\exp\{-cn^{1/5}\} \tag{A4}$$

in (A1).

For $-\delta \leqq \varphi \leqq \delta$ the integrand in (A1) can be rewritten

$$\begin{aligned}&\exp\{-ni\varphi + \exp(\beta e^{i\varphi}) - e^{\beta}\}\\ &= \exp\{-ni\varphi + \exp(\beta + i\beta\varphi - \tfrac{1}{2}\beta\varphi^2 - \tfrac{1}{6}i\beta\varphi^3 + 0(\beta\varphi^4)) - e^{\beta}\}\\ &= \exp\{-ni\varphi + e^{\beta}(1 + i\beta\varphi - \tfrac{1}{2}\beta\varphi^2 - \tfrac{1}{6}i\beta\varphi^3 - \tfrac{1}{24}i\beta^3\varphi^3\\ &\quad - \tfrac{1}{6}i\beta^3\varphi^3 + 0(\beta^4\varphi^4)) - e^{\beta}\}\end{aligned}$$

$$= \exp\{-\tfrac{1}{2}n\varphi^2(1+\beta)\} \times \{1 - \tfrac{1}{6}in(1+3\beta+\beta^2)\varphi^3$$

$$+ 0(n^2\beta^4\varphi^6 + n\beta^3\varphi^4)\} \tag{A5}$$

by (12), where the 0-notation refers to $n \to \infty$. Use has been made of $n\beta^3\varphi^3$ being small when $|\varphi| \leqq \delta$ and n is large; this follows from (12) and (A2).

The second term in (A5) is an odd function of φ, therefore its integral from $-\delta$ to $+\delta$ vanishes and we get

$$\int_{-\delta}^{\delta} = \int_{-\delta}^{\delta} \exp\{-\tfrac{1}{2}n\varphi^2(1+\beta)\}\, d\varphi$$

$$+ 0\left(\int_{-\infty}^{\infty} (n^2\beta^4\varphi^6 + n\beta^3\varphi^4) \exp\{-\tfrac{1}{2}n\varphi^2(1+\beta)\}\, d\varphi\right) \tag{A6}$$

$$= (\tfrac{1}{2}n(1+\beta))^{-1/2} \int_{-k}^{k} n^{-v^2}\, dv + 0(\beta^{1/2}n^{-3/2}),$$

where $k = \delta(\frac{1}{2}n(1+\beta))^{1/2}$. Now

$$\int_k^{\infty} e^{-v^2}\, dv < \int_k^{\infty} ve^{-v^2}\, dv = \tfrac{1}{2}e^{-k^2} = \tfrac{1}{2}\exp\{-\tfrac{1}{2}n(1+\beta)\delta^2\} < \tfrac{1}{2}\exp(-\tfrac{1}{2}n^{1/5}),$$

and a similar inequality holds for $\int_{-\infty}^{-k} e^{-v^2}\, dv$. Therefore replacement of the limits $\pm k$ by $\pm\infty$ in (A6) causes an error not exceeding $\exp(-\frac{1}{2}n^{1/5})$, and we get

$$\int_{-\delta}^{\delta} = (2\pi/n(1+\beta))^{1/2} + 0(\alpha^{1/2}n^{-3/2})$$

$$= (2\pi/n(1+\beta))^{1/2}\{1 + 0(\beta/n)\}. \tag{A7}$$

Summarizing (10), (11), (A1), (A3), (A4), and (A7), the leading term of (13) is obtained. The term with $0(\beta/n)$ (and if necessary, any further terms in the asymptotic expansion) can be obtained by carrying further the expansion under (A5).

5. Acknowledgment. We thank J. Riordan for bringing to our attention the isomorphism of distributions and equivalence relations, the availability of the unpublished table mentioned in footnote 1, and for the value of G_{51}.

REFERENCES

[1] ARFWEDSON, G., *Skand. Aktuarietids.*, Vol. 36 (1951), pp. 121–132.
[2] BELL, E. T., *Ann. Math.*, Vol. 39 (1938), pp. 539–557.
[3] EPSTEIN, L. F., *J. Math. Physics*, Vol. 18 (1939), pp. 153–173.
[4] FISHER, R. A., *Philos. Trans. Roy. Soc. London, Ser. A*, Vol. 222 (1922), pp. 309–369.
[5] HSU, L. C., *Ann. Math. Stat.*, Vol. 19 (1948), pp. 273–277.
[6] JEFFREYS, H., AND B. S. JEFFREYS, *Methods of Mathematical Physics*, University Press, Cambridge, 1946, p. 472.
[7] JORDAN, C., *Tôhoku Math J.*, Vol. 37 (1933), pp. 254–278; Vol. 38 (1953), p. 481.
[8] WALD, A., *Statistical Decision Functions*, John Wiley & Sons, New York, 1950.

ON TRANSIENT MARKOV CHAINS WITH APPLICATION TO THE UNIQUENESS PROBLEM FOR MARKOV PROCESSES[1]

By Leo Breiman

University of California, Berkeley

1. Summary. We focus our attention herein on a Markov chain $x_0, x_1, \cdots$ with a countable number of states indexed by a subset I of the integers and with stationary transition probabilities p_{ij}, and explore the sets of states defined by:

A transient set of states C is said to be *denumerably atomic* if $P(x_n \varepsilon C \text{ i.o.}) > 0$ and if for every infinite set $A \subset C$ we have $x_n \varepsilon C$ i.o. implies $x_n \varepsilon A$ i.o. with probability one (a.s.).

Following Blackwell's basic paper [1] which introduced the systematic use of martingales into the study of Markov chains, we use the semi-martingale convergence theorem [2] to characterize denumerably atomic sets in terms of the bounded solutions of the inequality

$$\phi(i) \leqq \sum_{j \varepsilon I} p_{ij}\phi(j), \qquad i \varepsilon I.$$

For chains whose state space contains a denumerably atomic set a convergence criterion for certain sums $\sum_{n=0}^{\infty} f(x_n)$ is then developed. The application of this criterion to a restricted class of continuous parameter Markov processes gives simple necessary and sufficient conditions for the existence of a unique process satisfying given infinitesimal conditions. This last result illuminates the connection between the necessary and sufficient conditions given by Feller [3] for uniqueness and the simpler conditions for birth and death processes given recently by Dobrusin [4], more recently by Karlin and McGregor [5], and by Reuter and Lederman [6] (see also [7]).

2. Characterization theorem.

Theorem 1. *The necessary and sufficient condition for a transient set of states C such that $P(x_n \varepsilon C \text{ i.o.}) > 0$ to be denumerably atomic is that any bounded solution $\phi(i)$ of*

$$\phi(i) \leqq \sum_{j \varepsilon I} p_{ij}\phi(j) \tag{A}$$

satisfy $\liminf_{i \varepsilon C} \phi(i) = \limsup_{i \varepsilon C} \phi(i)$.

Proof. Let C be denumerably atomic and $\phi(i)$ any bounded solution of (A). Then $E(\phi(x_n) \mid x_{n-1}, \cdots, x_0) \geqq \phi(x_{n-1})$ so that by the semi-martingale convergence theorem $\phi(x_n)$ converges a.s. to a function $f(\omega)$. Let A_1, A_2 be infinite subsets of C such that for $i \varepsilon A_1, \phi(i) < \alpha$ and for $i \varepsilon A_2, \phi(i) > \beta > \alpha$. Then, since almost every sample path which is in A_1 i.o. is in A_2 i.o. the limit of $\phi(x_n)$ cannot exist.

Received July 2, 1956.

[1] This paper was prepared with the support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171.

Conversely, let A be any infinite subject of C and take $\mathcal{E}$ to be the event that x_n is never in A. Let $\phi(i) = P(\mathcal{E} \mid x_0 = i)$, then:

$$\phi(i) = E(P(\mathcal{E} \mid x_0, x_1) \mid x_0 = i) = \sum_{j \in I-A} p_{ij}\phi(j) \leqq \sum_{j \in I} p_{ij}\phi(j).$$

Since $\phi(i)$ is zero on A, we have $\liminf_{i \in C} \phi(i) = 0$ whence $\limsup_{i \in C} \phi(i) = 0$. This implies that for almost every sample path $x_0, \cdots$ which is in C i.o., $\phi(x_n)$ converges to zero. But by the martingale convergence theorem, since $P(\mathcal{E} \mid x_n, \cdots, x_0) \leqq \phi(x_n)$, the indicator $I_{\mathcal{E}}(\omega)$ of the set $\mathcal{E}$ is zero for almost every ω ε [x_n ε C i.o.] and therefore almost every such ω is in A i.o.

We note, for future use, that if C is denumerably atomic and if $\phi(i)$ is a solution of (A) satisfying $E \mid \phi(x_n) \mid < K$, the conclusion $\liminf_{i \in C} \phi(i) = \limsup_{i \in C} \phi(i)$ remains unaltered.

It is interesting, as well as necessary, to know that any denumerably atomic set C can be embedded in a set $\bar{C}$ which is a maximal denumerably atomic set. That this is so follows from Blackwell's work [1] in the following sense: there is a set $\bar{C} \supset C$ such that $\bar{C}$ is denumerably atomic and x_n ε $\bar{C}$ i.o. implies x_n ε $\bar{C}$ for all sufficiently large n a.s.

3. Convergence criterion. The above characterization leads to a convergence criterion reminiscent of the Three-Series theorem.

THEOREM 2. *Let C be denumerably atomic and $f(i)$ a finite nonnegative function on I such that f is zero outside of C. Then the sum $\sum_{n=0}^{\infty} f(x_n)$ converges* a.s. *if and only if $\sum_{n=0}^{\infty} Ef(x_n)$ converges and otherwise diverges with probability equal to $P(x_n \varepsilon C$ i.o.).*

PROOF. Let $S = \sum_{n=0}^{\infty} f(x_n)$ and $\phi(i) = P(S < d \mid x_0 = i)$. Then $\phi(i)$ satisfies inequality (A). Suppose $\liminf_{i \in C} \phi(i) = 0$ for every value of d, then $\limsup_{i \in C} \phi(i) = 0$, and $\phi(x_n) \to 0$ a.s. on the set $[x_n \varepsilon C$ i.o.]. Since

$$P(S < d \mid x_n, \cdots, x_0) \leqq \phi(x_n)$$

it follows that $I_{[S<d]}(\omega) = 0$ a.s. for ω ε [x_n ε C i.o.] and hence that S diverges a.s. on this set.

From the above it follows that if S converges on some subset of [x_n ε C i.o.] of positive measure, there is a $\delta > 0$ and $d_1 > 0$ such that $P(S < d_1 \mid x_0 = i) \geqq \delta$ for i ε C. Let $R_m = \sum_{n=m}^{\infty} f(x_n)$ and define S_m as the set $[R_m < d_1]$. Writing

$$\int_{S_m} R_m - \int_{S_{m+1}} R_{m+1} = \int_{S_m} [R_m - R_{m+1}] - \int_{S_{m+1}-S_m} R_{m+1},$$

using the definition of S_m

$$\int_{S_m} f(x_m) \leqq d_1 P(S_{m+1} - S_m) + \int_{S_m} R_m - \int_{S_{m+1}} R_{m+1},$$

and summing over all m results in

$$\sum_{m=0}^{\infty} \int_{S_m} f(x_m) < 2\, d_1.$$

But,

$$\int_{S_m} f(x_m) = \sum_{i \varepsilon C} f(i) P(R_m < d_1 \mid x_m = i)\, P(x_m = i)$$

$$= \sum_{i \varepsilon C} f(i) P(S < d_1 \mid x_0 = i) P(x_m = i) \geqq \delta E f(x_m)$$

which proves the theorem.

COROLLARY 1. *Under the conditions of the above theorem, a necessary and sufficient condition for the* a.s. *convergence of* $\sum_{n=0}^{\infty} f(x_n)$ *is that the equation*

$$a(i) = f(i) + \sum_{j \varepsilon I} p_{ij} a(j) \tag{B}$$

have a bounded solution.

PROOF. Let $\sum_{n=0}^{\infty} f(x_n)$ converges a.s. to $S(\omega) < \infty$. Then $a(i) = E(S \mid x_0 = i)$ is a solution of (B) and $-a(i)$ is a solution of (A) with $Ea(x_n) \leqq ES$. If $a(i)$ is unbounded, then $\lim \sup_{i \varepsilon C} a(i) = \lim \inf_{i \varepsilon C} a(i) = \infty$, which implies that $a(x_n) \to \infty$ on a set of positive measure and contradicts the boundedness of $Ea(x_n)$. Conversely, if (B) has a bounded solution $a(i)$, then the iteration of (B) gives $\mid a(i) - E(\sum_{n=0}^{\infty} f(x_n) \mid x_0 = i) \mid \leqq \sup_{j \varepsilon I} a(j)$ which implies the convergence of $\sum_{n=0}^{\infty} Ef(x_n)$.

We relate Theorem 2 to the uniqueness problem which involves global structure, by confining ourselves to chains with a fairly simple decomposition. The following theorem is appropriate. Its proof follows immediately from the various definitions.

THEOREM 3. *Let the state space* I *of a Markov chain be completely decomposable into the set* C_0 *of recurrent states, a set* M *of transient states such that* $P(x_n \varepsilon M \text{ i.o.}) = 0$, *and a finite number of maximal denumerably atomic sets* $C_1, \cdots, C_N$. *If* f *is a function on* I *and if* f_k *is that function which equals* f *on* C_k *and is zero elsewhere, then* $\sum_{n=0}^{\infty} f(x_n)$ *diverges almost surely if and only if each* $\sum_{n=0}^{\infty} f_k(x_n)$ *diverges* a.s. *on the set* $[x_n \varepsilon C_k \text{ i.o.}]$.

4. The uniqueness problem. The synthetic uniqueness problem for continuous parameter Markov processes having states indexed by a subset I of the integers begins with a set of nonnegative constants q_i, p_{ij}, defined for $i, j \varepsilon I$, and asks concerning the existence of a unique process $X(t)$, $0 \leqq t < \infty$, having a given initial distribution and satisfying

(C)
i. $P(X(t)$ is constant in interval $[s, s + \tau] \mid X(s) = i) = 1 - q_i \tau + o(\tau)$
ii. P (First discontinuity of $X(t)$, $t \geqq s$, is a jump to $j \mid X(s) = i) = p_{ij}$.

Our remarks are restricted to the simple and common situation q_i, $p_{ij} < \infty$, $\sum_{j \varepsilon I} p_{ij} = 1$.

There is a general answer, [3], [6], and [8]: if no "explosions" are possible, if an infinite number of jumps cannot occur in any finite time interval, then there is a unique process satisfying (C) and having, in addition, all of the properties that could reasonably be desired. This is the "minimal" solution. In the contrary case,

there is in general no unique solution and the solutions that do exist are analytically or probabilistically pathological.

To be more exact; the traversal time of each path of infinite length $(i_1, i_2, \cdots)$ is a sum $Q_{i_1} + Q_{i_2} + \cdots$ of independent random variables with distributions $P(Q_{i_n} > t) = \exp(-q_{i_n}t)$. There is a Markov measure $\tilde{P}$ on the space of all paths induced by the p_{ij} and the given initial distribution. The minimal solution exists if and only if the transversal time is a.s. infinite for each path in a set of $\tilde{P}$ measure one. Using the Three-Series criterion for the divergence of a sum of independent random variables and writing $x_0, x_1, \cdots$ for the chain associated with the measure $\tilde{P}$ leads to an equivalent formulation.

Uniqueness criterion. A minimal solution exists if and only if $\sum_0^\infty 1/q_{x_n}$ diverges a.s.

The applicability of Theorem 2 is now apparent. For instance, in the birth and death process, the given constants are

if $i \geqq 1$, $q_i = \lambda_i + \mu_i$, $p_{i,i+1} = \lambda_i/(\lambda_i + \mu_i)$, $p_{i,i-1} = \mu_i/(\lambda_i + \mu_i)$, $p_{ij} = 0$ otherwise;

if $i = 0$, $q_i = 0$, $p_{00} = 1$, $p_{ij} = 0$ otherwise.

If return to the origin is uncertain, the positive integers form a maximal denumerably atomic set. The condition for the existence of a minimal solution, as given by Corollary 1, is that the equation

$$(\lambda_i + \mu_i)a(i) = 1 + \lambda_i a(i+1) + \mu_i a(i-1), \qquad i \geqq 1$$

have no bounded solution. A little formal computation yields the condition as stated in [4], [5], and [7].

Another interesting application is to the case

$$p_{ij} = p_{j-i}, \qquad 0 < \sum_{j \in I} j p_j < \infty,$$

where we restrict the state space I to those states with a positive probability of being entered. As pointed out to me by D. Blackwell, the basic theorem of renewal theory, Chung and Wolfowitz [9], provides the simplest proof that the nonnegative integers I^+ in I form a maximal denumerably atomic set. By this theorem, the expression

$$E(\text{number of entrances into } j \mid x_0 = i) = \sum_{n=0}^{\infty} p_{ij}^{(n)}$$

approaches a positive limit as $j \to +\infty$ through I, which implies that for any infinite set A of positive integers in I, $P(x_n \varepsilon A \text{ i.o.}) = 1$. As the negative integers I^- in I have the property $P(x_n \varepsilon I^- \text{ i.o.}) = 0$, the necessary and sufficient condition for the existence of the minimal solution is

$$\sum_{n=0}^{\infty} \sum_{j \varepsilon I^+} \frac{1}{q_j} p_{ij}^{(n)} = \infty.$$

Interchanging the order of summation, and applying the renewal theorem once more gives the equivalent condition

$$\sum_{j\neq 1^+} \frac{1}{q_j} = \infty.$$

A slight alteration of this discussion is sufficient to establish the same condition when the negative integers are absorbing states, that is, if

$$p_{ij} = p_{j-i} \quad \text{if} \quad i \geqq 0; \qquad q_i = 0 \quad \text{if} \quad i < 0.$$

If the state space of the chain $x_0, x_1, \cdots$ cannot be decomposed as indicated in Theorem 3, complications set in and Feller's criterion, which necessitates a close examination of every set of states A such that $P(x_n \varepsilon A \text{ all } n \mid x_0 \varepsilon A) > 0$ must be referred to. Simple necessary and sufficient conditions for uniqueness are possible only when some uniformity, such as denumerable atomicity, is present.

REFERENCES

[1] D. Blackwell, "On transient Markov processes, with a countable number of states and stationary transition probabilities," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 654–658.

[2] J. L. Doob, *Stochastic Processes*, Chapter 7, John Wiley and Sons, New York, 1953.

[3] W. Feller, "On the integro-differential equations of purely discontinuous Markov processes," *Trans. Amer. Math. Soc.*, Vol. 48 (1940), pp. 488–515.

[4] R. L. Dobrusin, "On the conditions of regularity of stationary Markov processes with a denumerable set of possible states," *Progr. Math. Sci. Moscow* (N.S.), Vol. 7 (1952), pp. 185–191.

[5] S. Karlin and J. McGregor, "Representation of a class of stochastic processes," *Proc. Nat. Acad. Sci.*, Vol. 41 (1955), pp. 387–391.

[6] W. Lederman and G. E. H. Reuter, "On the differential equations for the transition probabilities of Markov processes with enumerably many states," *Proc. Cambridge Philos. Soc.*, Vol. 49 (1953), pp. 247–262.

[7] M. S. Bartlett, *An Introduction to Stochastic Processes*, Cambridge, 1955.

[8] J. L. Doob, "Markov chains—denumerable case," *Trans. Amer. Math. Soc.*, Vol. 63 (1945), pp. 455–473.

[9] K. L. Chung and J. Wolfowitz, "On a limit theorem in renewal theory," *Ann. Mathematics*, Vol. 55 (1952), pp. 1–6.

AN APPROXIMATE FORMULA FOR THE CUMULATIVE z-DISTRIBUTION[1]

BY JOHN WISHART[2]

Statistical Laboratory, University of Cambridge and Princeton University

1. Summary. A straightforward expansion and integration of the frequency function for Fisher's z produces a formula for the probability that z is not exceeded, of which the successive terms decrease rapidly when n_1 and n_2 are large. It is given in terms of incomplete normal moment functions (or χ^2 probabilities), and as a polynomial in $zN^{1/2}$, where N is the harmonic mean of n_1 and n_2. This last form is identical with the inverted Cornish-Fisher expansion, originally deduced by quite different methods.

2. To obtain their well-known expansion for determining percentage points for the distribution of z (one-half of the natural logarithm of the ratio of two independent variance estimates from normal data) in cases where the degrees of freedom n_1 and n_2 are large, Cornish and Fisher (1937) used the method of the normalizing transformation. They developed a Gram-Charlier Type A series expansion which required knowledge of the cumulants of z. These they worked out in the approximate form for large n_1 and n_2, to a point sufficient for the order of approximation worked to. The method is rather complicated, but a final formula is given which enables chosen percentage points to be determined. Although it is possible by substitution to deduce the corresponding formula for determining the probability associated with a chosen value of z, the author does not recall having seen such a formula explicitly stated.[3]

3. The frequency function of z may be manipulated directly so as to give on integration this inverted formula. The method is direct and simple, requires no Gram-Charlier Type A series, and no cumulants.

Consider two independent variance estimates s_1^2 and s_2^2 from normal data, having degrees of freedom n_1 and n_2. z is then $\frac{1}{2} \ln (s_1^2 / s_2^2)$. For the time being write $\frac{1}{2}n_1$ as c_1 and $\frac{1}{2}n_2$ as c_2. Then the frequency function of z is

$$\frac{2(c_1/c_2)^{c_1}}{\mathrm{B}(c_1, c_2)} \frac{e^{2c_1 z}}{(1 + c_1 e^{2z}/c_2)^{c_1+c_2}}, \tag{1}$$

where the range of z is from $-\infty$ to ∞, and $\mathrm{B}(c_1, c_2)$ is the Beta-Function, equal

Received February 22, 1956.

[1] Research partially supported by the Office of Naval Research.

[2] This paper was recommended for publication after the death of the author. It is published, with minor emendations, after consulting with a colleague of the author. The original title was "A new derivation of the inverted Cornish-Fisher expansion for the z-distribution." Ed.

[3] Reviewers note. Campbell [1] gave an expression for finding percentage points; he did not require a knowledge of cumulants or use the Gram-Charlier Type A series expansion. Student [4] used such an expansion for t to compute his original table, and Fisher [3] developed the expansion by methods similar to those used here.

to $\Gamma(c_1)\Gamma(c_2) / \Gamma(c_1 + c_2)$. We shall take n_1 as the smaller of the degrees of freedom, so that $c_1 / c_2 \leqq 1$. The frequency function may be written

$$\frac{2(c_1/c_2)^{c_1}}{B(c_1, c_2)} \exp\{2c_1 z - (c_1 + c_2) \ln (1 + c_1 e^{2z}/c_2)\}$$

$$= \frac{2(c_1/c_2)^{c_1}}{B(c_1, c_2)} \exp\left[2c_1 z - (c_1 + c_2) \ln \frac{c_1 + c_2}{c_2}\right. \tag{2}$$

$$\left. - (c_1 + c_2) \ln\left\{1 + \frac{c_1(e^{2z} - 1)}{c_1 + c_2}\right\}\right].$$

The first logarithm can be put into the outside term, and the second may be expanded, noting that $c_1(e^{2z} - 1) / (c_1 + c_2)$ will lie between $+1$ and -1 except in the extreme tail of the distribution when n_1 and n_2 are nearly equal and of the order of 30 or less.

The frequency function then becomes

$$\frac{2c_1^{c_1}c_2^{c_2}}{(c_1 + c_2)^{c_1+c_2}B(c_1, c_2)} \exp\left[2c_1 z - c_1(2z) - \frac{C}{2} \cdot \frac{(2z)^2}{2!} - \frac{C}{2}\frac{c_2 - c_1}{c_1 + c_2}\frac{(2z)^3}{3!}\right.$$

$$- \frac{C}{2}\left(1 - \frac{3C}{c_1 + c_2}\right)\frac{(2z)^4}{4!} - \frac{C}{2}\frac{c_2 - c_1}{c_1 + c_2}\left(1 - \frac{6C}{c_1 + c_2}\right)\frac{(2z)^5}{5!} \tag{3}$$

$$\left. - \frac{C}{2}\left(1 - \frac{15C}{c_1 + c_2} + \frac{30C^2}{(c_1 + c_2)^2}\right)\frac{(2z)^6}{6!} - \cdots\right].$$

where C is the harmonic mean of c_1 and $c_2 = 2\,c_1c_2 / (c_1 + c_2)$.

Now put $2z = x(2/C)^{1/2}$, whereupon the frequency function may be written

$$\frac{\sqrt{(2\pi)}c_1^{c_1-1/2}c_2^{c_2-1/2}}{(c_1 + c_2)^{c_1+c_2-1/2}B(c_1, c_2)} \cdot \frac{e^{-x^2/2}}{\sqrt{(2\pi)}}$$

$$\cdot \exp\left\{-\left[\left(\frac{2}{C}\right)^{0.5} \cdot \frac{c_2 - c_1}{c_1 + c_2} \cdot \frac{x^3}{3!} + \frac{2}{C}\left(1 - \frac{3C}{c_1 + c_2}\right)\frac{x^4}{4!} + \left(\frac{2}{C}\right)^{1.5}\right.\right. \tag{4}$$

$$\left.\left. \cdot \frac{c_2 - c_1}{c_1 + c_2}\left(1 - \frac{6C}{c_1 + c_2}\right)\frac{x^5}{5!} + \left(\frac{2}{C}\right)^2\left(1 - \frac{15C}{c_1 + c_2} + \frac{30C^2}{(c_1 + c_2)^2}\right)\frac{x^6}{6!} + \cdots\right]\right\}.$$

On expanding the Γ-functions in $B(c_1, c_2)$ by Stirling's formula, the first part of this expression becomes approximately

$$\left(1 + \frac{1}{12(c_1 + c_2)} + \frac{1}{288(c_1 + c_2)^2}\right)\left(1 + \frac{1}{12c_1} + \frac{1}{288c_1^2}\right)^{-1}$$

$$\cdot \left(1 + \frac{1}{12c_2} + \frac{1}{288c_2^2}\right)^{-1} = 1 - \frac{1}{6N}\left(2 - \frac{N}{n_1 + n_2}\right) \tag{5}$$

$$+ \frac{1}{72N^2}\left(2 - \frac{N}{n_1 + n_2}\right)^2$$

in terms of n_1 and n_2 and their harmonic mean N.

The second part of (4) is the normal frequency function, and the third part may be expanded into the following series in terms of n_1, n_2 and N:

$$
\begin{aligned}
&1 - \frac{1}{3N^{0.5}} \cdot \frac{n_2 - n_1}{n_1 + n_2} x^3 + \frac{1}{18N}\left[\left(1 - \frac{2N}{n_1 + n_2}\right)x^6 - 3\left(1 - \frac{3N}{n_1 + n_2}\right)x^4\right] \\
&\quad - \frac{1}{810N^{1.5}} \cdot \frac{n_2 - n_1}{n_1 + n_2}\left[5\left(1 - \frac{2N}{n_1 + n_2}\right)x^9 - 45\left(1 - \frac{3N}{n_1 + n_2}\right)x^7\right. \\
(6) \qquad &\quad \left. + 54\left(1 - \frac{6N}{n_1 + n_2}\right)x^5\right] + \frac{1}{9720N^2}\left[5\left(1 - \frac{2N}{n_1 + n_2}\right)^2 x^{12}\right. \\
&\quad \left. - 90\left(1 - \frac{2N}{n_1 + n_2}\right)\left(1 - \frac{3N}{n_1 + n_2}\right)x^{10}\right. \\
&\left. + 27\left(13 - \frac{94N}{n_1 + n_2} + \frac{141N^2}{(n_1 + n_2)^2}\right)x^8 - 216\left(1 - \frac{15N}{n_1 + n_2} + \frac{30N^2}{(n_1 + n_2)^2}\right)x^6\right]
\end{aligned}
$$

as far as terms in N^{-2}.

We now have a frequency function for the variable $x = zN^{1/2}$ $(-\infty \leqq x \leqq \infty)$ in terms of the normal frequency function multiplied by a polynomial in x. For a chosen X, the probability $P(x \leqq X)$ is given by the integral of the frequency function from $-\infty$ to X. Alternative forms can be found for the result of the integration. We may express it in terms of Pearson's incomplete normal moment functions

$$\mu_r(X) = \frac{1}{\sqrt{(2\pi)}} \int_0^X x^r e^{-x^2/2}\, dx$$

$$m_r(X) = \frac{\mu_r(X)}{(r-1)(r-3)\cdots 1 \text{ or } 2}$$

according as r is even or odd. Numerical values for $m_{12}(X)$ are given to seven decimals in *Tables for Statisticians and Biometricians* (Pearson, 1914, 1931), while $\mu_0(X) = P(X) - 0.5$, where

$$P(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^X e^{-x^2/2}\, dx$$

is given elsewhere in the same *Tables* (Table II), and also by Pearson and Hartley (1954), Table 1.

In this form the probability $P(0 \leqq x \leqq X)$ is

$$
\begin{aligned}
&\left[1 - \frac{1}{6N}\left(2 - \frac{N}{n_1 + n_2}\right) + \frac{1}{72N^2}\left(2 - \frac{N}{n_1 + n_2}\right)^2\right] \\
&\qquad \cdot \left\{\mu_0(X) - \frac{2}{3N^{0.5}} \cdot \frac{n_2 - n_1}{n_1 + n_2} m_3(X)\right. \\
&\quad + \frac{1}{6N}\left[5\left(1 - \frac{2N}{n_1 + n_2}\right)m_6(X) - 3\left(1 - \frac{3N}{n_1 + n_2}\right)m_4(X)\right]
\end{aligned}
$$

$$
\begin{aligned}
(7)\quad & -\frac{8}{135N^{1.5}}\cdot\frac{n_2-n_1}{n_1+n_2}\left[40\left(1-\frac{2N}{n_1+n_2}\right)m_9(X)\right.\\
& -45\left(1-\frac{3N}{n_1+n_2}\right)m_7(X)+9\left(1-\frac{6N}{n_1+n_2}\right)m_5(X)\Big]\\
& +\frac{1}{72N^2}\left[385\left(1-\frac{2N}{n_1+n_2}\right)^2 m_{12}(X)-630\left(1-\frac{2N}{n_1+n_2}\right)\right.\\
& \cdot\left(1-\frac{3N}{n_1+n_2}\right)m_{10}(X)+21\left(15-\frac{94N}{n_1+n_2}+\frac{141N^2}{(n_1+n_2)^2}\right)m_8(X)\\
& \left.-24\left(1-\frac{15N}{n_1+n_2}+\frac{30N^2}{(n_1+n_2)^2}\right)m_6(X)\right]\Big\}.
\end{aligned}
$$

The probability $P(-\infty \leqq x \leqq 0)$ is got from (7) as a special case by putting $\mu_0(-\infty) = m_{2r}(-\infty) = 0.5$, and $m_{2r+1}(-\infty) = (-2\pi)^{-1/2}$. It then becomes[4]

$$
\begin{aligned}
(8)\quad & \left[1-\frac{1}{6N}\left(2-\frac{N}{n_1+n_2}\right)+\frac{1}{72N^2}\left(2-\frac{N}{n_1+n_2}\right)^2\right]\\
& \cdot\left\{\frac{1}{2}+\frac{1}{12N}\left(2-\frac{N}{n_1+n_2}\right)+\frac{1}{144N^2}\left(2-\frac{N}{n_1+n_2}\right)^2\right.\\
& \left.+\frac{2}{3\sqrt{(2\pi N)}}\cdot\frac{n_2-n_1}{n_1+n_2}\left[1+\frac{4}{45N}\left(4+\frac{N}{n_1+n_2}\right)\right]\right\}\\
& =\frac{1}{2}+\frac{2}{3\sqrt{(2\pi N)}}\cdot\frac{n_2-n_1}{n_1+n_2}\left[1+\frac{1}{90N}\left(2+\frac{23N}{n_1+n_2}\right)\right].
\end{aligned}
$$

The sum or difference of (7) and (8), according as X is positive or negative, gives the probability $P(-\infty \leqq x \leqq X)$.

Alternatively we may write in (7)

$$(2\pi)^{\frac{1}{2}}m_{2r+1}(X) = P(X^2 \mid 2r+2),$$

$$2m_{2r}(X) = P(X^2 \mid 2r+1),$$

where $P(X^2 \mid \nu)$ denotes the probability that χ^2 does not exceed X^2, for ν degrees of freedom. These probabilities may be obtained to five decimals by subtracting from unity the χ^2 probabilities given in Pearson and Hartley (1954), Table 7.

A series expansion for the probability $P(-\infty \leqq x \leqq X)$ can be obtained in terms of $P(X)$ and $Z(X) = e^{-x^2/2} / \nu(2\pi)$, together with a polynomial in X, by associating (6) with $Z(x)$ and integrating term by term by parts. This gives the required probability as

$$\left\{1-\frac{1}{6N}\left(2-\frac{N}{n_1+n_2}\right)+\frac{1}{72N^2}\left(2-\frac{N}{n_1+n_2}\right)^2\right\}$$

[4] Reviewer's note. The algebraic signs for $m_{2r}(-\infty)$ should be the opposites of those given here; when X is negative, (7) takes negative values.

$$\cdot\left[P(X)+\frac{n_2-n_1}{n_1+n_2}\cdot\frac{X^2+2}{3N^{0.5}}Z(X)-\frac{1}{18N}\left\{\left(1-\frac{2N}{n_1+n_2}\right)X^5\right.\right.$$
$$\left.+\left(2-\frac{N}{n_1+n_2}\right)(X^3+3X)\right\}Z(X)+\frac{1}{6N}\left(2-\frac{N}{n_1+n_2}\right)P(X)$$
$$+\frac{1}{810N^{1.5}}\frac{n_2-n_1}{n_1+n_2}\left\{5\left(1-\frac{2N}{n_1+n_2}\right)X^9-5\left(1-\frac{11N}{n_1+n_2}\right)X^6\right.$$
$$\left.+6\left(4+\frac{N}{n_1+n_2}\right)(X^4+4X^2+8)\right\}Z(X)$$
$$-\frac{1}{9720N^2}\left\{5\left(1-\frac{2N}{n_1+n_2}\right)^2X^{11}-5\left(1-\frac{2N}{n_1+n_2}\right)\right.$$
$$\cdot\left(7-\frac{32N}{n_1+n_2}\right)X^9+9\left(4-\frac{52N}{n_1+n_2}+\frac{103N^2}{(n_1+n_2)^2}\right)X^7$$
$$\left.+9\left(2-\frac{N}{n_1+n_2}\right)^2(X^5+5X^3+15X)\right\}Z(X)$$
$$\left.+\frac{1}{72N^2}\left(2-\frac{N}{n_1+n_2}\right)^2P(X)\right].$$

On multiplying in by the outside factor this becomes

$$P(X)+Z(X)\left[\frac{n_2-n_1}{n_1+n_2}\cdot\frac{X^2+2}{3N^{0.5}}-\frac{1}{18N}\left\{\left(1-\frac{2N}{n_1+n_2}\right)X^5\right.\right.$$
$$\left.+\left(2-\frac{N}{n_1+n_2}\right)(X^3+3X)\right\}+\frac{1}{810N^{1.5}}\frac{n_2-n_1}{n_1+n_2}\left\{5\left(1-\frac{2N}{n_1+n_2}\right)X^9\right.$$
$$-5\left(1-\frac{11N}{n_1+n_2}\right)X^6+6\left(4+\frac{N}{n_1+n_2}\right)X^4$$
$$\left.+3\left(2+\frac{23N}{n_1+n_2}\right)\left(X^2+2\right)\right\}-\frac{1}{9720N^2}\left\{5\left(1-\frac{2N}{n_1+n_2}\right)^2X^{11}\right. \tag{9}$$
$$-5\left(1-\frac{2N}{n_1+n_2}\right)\left(7-\frac{32N}{n_1+n_2}\right)X^9+9\left(4-\frac{52N}{n_1+n_2}\right.$$
$$\left.+\frac{103N^2}{(n_1+n_2)^2}\right)X^7-9\left(2-\frac{N}{n_1+n_2}\right)\left(8-\frac{19N}{n_1+n_2}\right)X^5$$
$$\left.-45\left(2-\frac{N}{n_1+n_2}\right)^2(X^3+3X)\right\}\Bigg].$$

This is the expression which is the "direct" form of the Cornish-Fisher expansion, yielding, to terms in N^{-2}, the probability that z shall not exceed $xN^{-1/2}$. Additional terms could be worked out by noting that the terms of the exponential in (3) are equivalent to the binomial cumulants, but the terms in (9) should

suffice for W of the order of 50 or above, and fewer terms will do if the probability is not required to a large number of significant figures.

For the benefit of those accustomed to the notation of Cornish and Fisher, (9) may be put into the form

$$\begin{aligned}\mu + z\Bigg[&\frac{\sqrt{(\tfrac{1}{2}\sigma)}}{3}\cdot\frac{\delta}{\sigma}(X^2+2) - \frac{\tfrac{1}{2}\sigma}{36}\left\{3(X^3+3X) + \frac{\delta^2}{\sigma^2}(2X^3+X^2+3X)\right\}\\ &+\frac{(\tfrac{1}{2}\sigma)^{1.5}}{1620}\cdot\frac{\delta}{\sigma}\left\{9(5X^6+6X^4+9X^2+18) + \frac{\delta^2}{\sigma^2}\right.\\ &\left.\cdot(10X^8-55X^6-6X^4-69X^2-138)\right\} - \frac{(\tfrac{1}{2}\sigma)^2}{38880}\\ &\cdot\left\{27(5X^7+3X^5-15X^3-45X) + 18\frac{\delta^2}{\sigma^2}\right.\\ &\cdot(10X^9-51X^7-27X^5-15X^3-45X)\\ &\left.+\frac{\delta^4}{\sigma^4}(20X^{11}-320X^9+927X^7-171X^5-45X^3-135X)\right\}\Bigg],\end{aligned} \tag{10}$$

but note that z here is what we have hitherto written as $Z(X)$. In using (10) we take X as the chosen value of Fisher's z divided by $\sqrt{(\frac{1}{2}\sigma)}$, i.e. by its approximate standard derivation, $[\frac{1}{2}(1/n_1 + 1/n_2)]^{1/2}$; δ is, of course, $1/n_1 - 1/n_2$.

The order of the terms in (9) or (10) may be seen if we choose as an example $n_1 = 60, n_2 = 120, Z = (\sqrt{5})/20 = 0.1118034$. Then $N = 80$, while $\sqrt{(\frac{1}{2}\sigma)} = (\sqrt{5})/20$ and $\delta/\sigma = \frac{1}{3}$, and $X = 1$. Using Pearson and Hartley's Table 1 we find for the probability that this chosen value of Z is not exceeded

	0.8413	447
+	90	177
−	10	642
+		218
+		16
	0.8493	216

so that we are here close to the 15 per cent point of the z distribution.

When $n_1 = n_2 = n$ we have $N = n$, $N/(n_1 + n_2) = \frac{1}{2}$, also $\delta = 0$, $\frac{1}{2}\sigma = n^{-1}$. Then (7) and (8) give

$$\tfrac{1}{2} + \left(1 - \frac{1}{4n} + \frac{1}{32n^2}\right)\left(\mu_0(X) + \frac{m_4(X)}{4n} - \frac{32m_6(X) - 35m_8(X)}{96n^2}\right), \tag{11}$$

while (9) or (10) gives

$$P(X) - Z(X)\left\{\frac{X(X^2+3)}{12n} + \frac{5X^7+3X^5-15X(X^2+3)}{1440n^2}\right\}. \tag{12}$$

When $n_1 = n$, $n_2 = \infty$, we have an expansion from which can be calculated the probability that a chosen value of χ^2, for n degrees of freedom, is not exceeded.

If this value be denoted χ_0^2, then we take $X = \sqrt{(\frac{1}{2}n)} \cdot \ln(\chi_0^2/n)$, so that we are effectively transforming χ^2 by first forming the ratio of χ^2 to its mean, raised to the power of its standard deviation, and then taking one-half the natural logarithm of this quantity. The expansion for the probability may be obtained from (7) and (8), or from (9), by putting $N = 2n$, $(n_2 - n_1)/(n_1 + n_2) = 1$ and $N/(n_1 + n_2) = 0$, or from (10) with $\sigma = \delta = n^{-1}$. It has been developed from first principles by the author in [7].

REFERENCES

[1] G. A. CAMPBELL, "Probability curves showing Poisson's exponential summation," *Bell System Technical Journal*, Vol. 2 (1923), pp. 95–113; and Collected Papers, N. Y. (1937), pp. 224–242.

[2] E. A. CORNISH AND R. A. FISHER, *Revue de l'Institut International de Statistique*, Vol. 4 (1937), p. 307.

[3] R. A. FISHER, "The asymptotic approach to Behrens' integral with further tables for the d-test of significance," *Ann. Eng. Lond.*, Vol. 11 (1941), p. 151.

[4] M. G. KENDALL, "Advanced Theory of Statistics," Vol. II, Section 21.10, p. 101.

[5] K. PEARSON, *Tables for Statisticians and Biometricians*, Part I (Table IX); Part II (Table XIII), Biometric Laboratory, London, 1914, 1931.

[6] E. S. PEARSON AND H. O. HARTLEY, *Biometrika Tables for Statisticians*, Vol. I, Cambridge University Press, 1954.

[7] J. WISHART, "χ^2 probabilities for large numbers of degrees of freedom" *Biometrika*, Vol. 43 (1956), pp. 92–95.

THE MIXTURE OF NORMAL DISTRIBUTIONS WITH DIFFERENT VARIANCES[1]

BY D. TEICHROEW[2]

University of California, Los Angeles

1. Introduction. In some practical problems, the observed variable may have a normal distribution whose variance varies from one observation to the next. The purpose of this note is to give the formula for the marginal distribution when the variances are assumed to be distributed according to the Gamma distribution.

2. The distribution in the general case. We assume that the conditional density of X, given σ^2, is

$$f(x/\sigma^2) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} e^{-x^2/2\sigma^2} \qquad -\infty < x < \infty, \qquad \sigma^2 > 0,$$

Received May 25, 1956; revised July 18, 1956.

[1] The preparation of this paper was sponsored (in part) by the Office of Naval Research, USN.

[2] Present address: National Cash Register Company, Hawthorne, California.

and that the density function of the variance is

$$g(\sigma^2) = \frac{\alpha^\lambda}{\Gamma(\lambda)} e^{-\alpha\sigma^2}(\sigma^2)^{\lambda-1} \qquad \alpha > 0, \qquad \lambda > 0.$$

Multiplying these two densities together and integrating immediately yields the marginal density function of X in the form

$$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)(2\pi)^{1/2}} \int_0^\infty \exp\{-[\alpha\sigma^2 + (x^2/2\sigma^2)]\}\ (\sigma^2)^{\lambda-3/2}\, d\sigma^2,$$

which, using a formula for the modified Hankel function [3], p. 39, gives

$$f(x) = \frac{\alpha^{1/2}(x\sqrt{2\alpha})^{\lambda-1/2}k_{\lambda-1/2}(x\sqrt{2\alpha})}{\sqrt{\pi}2^{\lambda-1}\Gamma(\lambda)}.$$

The distribution function of X could be obtained by integrating the density function or by evaluating two hypergeometric functions, for, by the Paul Lévy inversion formula ([4], p. 93, Eq. (10.3.1)) the well-known relation between $\sin x$ and $J_{1/2}(x)$, and Formula 1 of [2] (p. 434), we have

$$F(x) = \tfrac{1}{2} + \frac{2(2\alpha x^2)}{\sqrt{2\pi}}\left[\frac{\Gamma(\frac{1}{2})\Gamma(\lambda - \frac{1}{2})}{(2\alpha x^2)^{\lambda-1/2}2^{3/2}\Gamma(\lambda)\Gamma(\frac{3}{2})}\ {}_1F_2\left(\frac{1}{2}, \frac{3}{2} - \lambda, \frac{3}{2}, \frac{\alpha x^2}{2}\right)\right.$$
$$\left. + \frac{\Gamma(\frac{1}{2} - \lambda)}{2^{2\lambda+1/2}\Gamma(\lambda+1)}\ {}_1F_2\left(\lambda, \lambda+1, \lambda+\frac{1}{2}, \frac{\alpha x^2}{2}\right)\right],$$

where ${}_1F_2$ denotes a generalized hypergeometric function defined as

$${}_1F_2(\beta_1, \gamma_1, \gamma_2; z) = \sum_{n=0}^{\infty} \frac{(\beta_1)_n}{(\gamma_1)_n(\gamma_2)_n} z^n,$$

where $(\beta)_n = \beta(\beta+1)\cdots(\beta+n-1)$; $(\beta)_0 = 1$.

The density and distribution function can also be obtained from the characteristic function which is

$$\phi(t) = \frac{1}{(1 + t^2/2\alpha)^\lambda}.$$

3. The distribution when λ is an integer. For $\lambda = n$, an integer, from [1], p. 40 and [1], p. 128, No. 67b, we get

$$f(x) = \frac{\sqrt{2\alpha}}{(n-1)!}\frac{e^{-\alpha x\sqrt{2}}}{2^{2n-1}}\sum_{v=0}^{n-1}\frac{(2n-v\,2)!(2\cdot\alpha x\sqrt{2})^v}{v!(n-z-1)!}.$$

The distribution function can also be expressed in closed form if $\lambda = n$ an integer by the following formula ([1], p. 127, No. 66c)

$$\int_0^\infty \frac{\sin xt}{(a^2+t^2)^n}\frac{dt}{t} = \frac{\pi}{2a^{2n}}\left[1 - \frac{e^{-ax}}{2^{n-1}(n-1)!}F_{n-1}(ax)\right],$$

where $F_0(z) = 1$, $F_1(z) = z + 2$, and $F_n(z) = (z + 2n)\ F_{n-1}(z) - zF'_{n-1}(z)$, for $a > 0; x \geqq 0; n = 1, 2, 3 \cdots$. These recurrence relations could be used to compute a table of the distribution function.

4. Moments. The moments are obtainable directly from the expansion of the characteristic function

$$\frac{1}{\left(1 + \frac{t^2}{2\alpha}\right)^{\lambda}} = 1 - \frac{\lambda}{\alpha}\frac{t^2}{2} + \frac{\lambda(\lambda+1)}{\alpha^2}\frac{t^4}{2!4} - \frac{\lambda(\lambda+1)(\lambda+2)}{\alpha^3}\frac{t^6}{3!8}.$$

We have

$$\mu_1' = 0 \qquad 0 = \mu_2' = \mu_3' = \mu_7' = \cdots$$

$$\mu_2 = \mu_2' = \frac{\lambda}{\alpha}$$

$$\mu_4 = \frac{3\lambda(\lambda+1)}{\alpha^2}$$

$$\beta_1 = 0, \beta_2 = \frac{\mu_4}{\mu_2^2} = 3\left(1 + \frac{1}{\lambda}\right).$$

As one would expect, the variance of X increases as λ increases. It is interesting to note that β_2 is always greater than 3.

REFERENCES

1. W. GRÖBNER AND N. HOFREITER, *Integraltafel, Zweiter Teil, Bestimmte Integrale*, Springer-Verlag 1950.
2. G. N. WATSON, *A Treatise on the Theory of Bessel Functions*, MacMillan, New York, 2nd edition, 1948.
3. W. MAGNUS AND F. OBERHETTINGER, *Formeln and Sätze für die Speziellen Funktionen der Mathematischen Physik*, Springer Verlag, Berlin, 1948.
4. H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

METRICS AND NORMS ON SPACES OF RANDOM VARIABLES

BY A. J. THOMASIAN[1]

University of California, Berkeley

1. Introduction and summary. Let $\mathfrak{X}$ be the space of random variables defined on an abstract probability space $(\Omega, \mathfrak{A}, P)$ where we consider any two elements of $\mathfrak{X}$ which are equal a.s. (almost surely) as the same. Fréchet [2] exhibited a metric on $\mathfrak{X}$ (for example, $E[|X - Y|/(1 + |X - Y|)]$) with the property that con-

Received May 23, 1956; revised October 8, 1956.

[1] This paper was prepared while the author held a National Science Foundation Fellowship.

vergence in the metric is equivalent to convergence in probability, and he showed that for some probability spaces the same cannot be done for convergence a.s. Dugué [1] showed that it is not in general possible to define a norm on $\mathfrak{X}$ such that convergence in the norm is equivalent to convergence in probability. These results are contained in and completed by the following fact which was stated without proof by the author in [5] and which follows easily from the two theorems stated and proved in this note. There exists a metric (norm) on $\mathfrak{X}$ with convergence in the metric (norm) equivalent to convergence a.s. (in probability) if, and only if, Ω is the union of countable (finite) number of disjoint atoms. After these results were obtained it was found that the equivalence of parts (ii) and (iii) of Theorem 1 had been proved by Marczewski [4], p. 121.

An atom of a probability space is a measurable set A with $P(A) > 0$, such that any measurable subset has probability 0 or $P(A)$. It is easy to show that a random variable is a.s. constant on an atom. f will always designate a real-valued function defined on $\mathfrak{X}$. Convergence in f is said to be equivalent to convergence a.s. (in probability) if, for every sequence $\{X_n\}$ of elements from $\mathfrak{X}$, $f(X_n) \to 0$ if, and only if, $X_n \to 0$ a.s. (in probability).

THEOREM 1. *The following conditions on a probability space are equivalent.*

(i) *There exists a function f, such that convergence in f is equivalent to convergence a.s.*

(ii) *For any sequence $\{X_n\}$ from $\mathfrak{X}$, if $X_n \to 0$ in probability, then $X_n \to 0$ a.s.*

(iii) *Ω is a countable union of disjoint atoms.*

THEOREM 2. *The following conditions on a probability space are equivalent.*

(a) *There exists a function f, such that convergence in f is equivalent to convergence in probability and f satisfies $|f(\alpha X)| = |\alpha| \cdot |f(X)|$ for any $X \in \mathfrak{X}$ and any real number α.*

(b) *Ω is a finite union of disjoint atoms.*

2. Proof of Theorem 1. The following well-known result (see Loève [3], p. 100, Example 7) will be used in the proof.

THEOREM A. *For any probability space, $\Omega = A + \sum_1^\infty A_i$ where all of the sets in the decomposition are disjoint and each A_i is the empty set or an atom, and for every measurable subset B of A, P takes every value between 0 and $P(B)$ for measurable subsets of B.*

(ii) implies (i) by the result of Fréchet.

To show that (i) implies (ii) assume (i) and take any sequence $X_n \to 0$ in probability. If $f(X_n) \not\to 0$ then there exists a subsequence $X_{n'}$, and an $\varepsilon > 0$ such that $|f(X_{n'})| > \epsilon$. But $X_{n'} \to 0$ in probability so that it has a subsequence $X_{n''} \to 0$ a.s. Thus $f(X_{n''}) \to 0$ contradicting $|f(X_{n'})| > \epsilon$. Therefore, $f(X_n)$ must converge to 0, hence, $X_n \to 0$ a.s.

(ii) follows easily from (iii) since a random variable is a.s. constant on an atom.

To prove that (ii) implies (iii), assume that (iii) is false. Thus in the decomposition of Theorem A, $P(A) > 0$ and for each n, $A = \sum_{i=1}^n A_{ni}$ where $P(A_{ni}) = (1/n)P(A)$ for $i = 1, 2, \cdots, n$, and the sets $A_{n1}, A_{n2}, \cdots, A_{nn}$ are disjoint.

Let X_{ni} be the characteristic function of the set A_{ni}. The sequence of random variables

$$X_{11}, X_{21}, X_{22}, X_{31}, \cdots$$

converges to 0 in probability but not a.s. so that (ii) implies (iii), completing the proof.

3. Proof of Theorem 2. To prove that (a) implies (b), assume that (a) is true and (b) is false. From Theorem A there exists a sequence A_n of events with $0 < P(A_n) \to 0$. Let X_n be the characteristic function of the set A_n. For all n, $f(X_n) \neq 0$ because if $f(X_{n_0}) = 0$, then by (a) the sequence of random variables, each of which is X_{n_0}, must converge to 0 in probability, contradicting $P(A_{n_0}) > 0$. By (a), $[f(X_n/f(X_n))] = 1$ for all n, so that the sequence of random variables $X_n/f(X_n)$ cannot converge to 0 in probability. However, it must, because $P(A_n) \to 0$. A contradiction has been reached, hence (a) implies (b).

Assuming (b) it is easy to show that $f(X) = E \mid X \mid$ is a norm on $\mathfrak{X}$ such that convergence in f is equivalent to convergence in probability. Theorem 2 is proved.

4. Acknowledgment. The author wishes to thank Professor M. Loève for suggesting this problem.

REFERENCES

[1] D. Dugué, "L'existence d'une norme est incompatible avec la convergence en probabilité," *C. R. Acad. Sci., Paris*, Vol. 240 (1955), p. 1307.
[2] M. Fréchet, *"Généralites sur les Probabilités. Elements Aléatoires*, Gauthier-Villars, 1935.
[3] M. Loève, *Probability Theory*, D. Van Nostrand, New York, 1955.
[4] E. Marczewski, "Remarks on the convergence of measurable sets and measurable functions," *Colloquium Math.*, Vol. 3 (1955), pp. 118–124.
[5] A. J. Thomasian, "Distances et normes sur les espaces de variables aléatoires," *C. R. Acad. Sci., Paris*, Vol. 242 (1956), p. 447.

DIVERGENT TIME HOMOGENEOUS BIRTH AND DEATH PROCESSES[1]

By Peter W. M. John

University of New Mexico

1. Introduction. In a time-homogeneous birth and death process a population is considered, the size of which is given by the random variable $n(t)$ defined on the non-negative integers. If at time t the population size is n, the probability that a birth occurs in the time interval $(t, t + \Delta t)$ is $\lambda_n t + o(\Delta t)$; the probability of a death is $\mu_n t + o(\Delta t)$, and the probability of the occurrence of more than one

Received January 17, 1956; revised September 24, 1956.

[1] These results were included in a dissertation submitted to the University of Oklahoma in partial fulfillment of the requirements for the Ph.D. degree in mathematics, August, 1955.

event is $o(\Delta t)$. The parameters λ_n and μ_n are non-negative and are independent of t. The probabilities $p_n(t)$ that the population size is n at time t then satisfy the inequality, Feller [4], $\sum_0^\infty p_n(t) \leqq 1$. We shall impose the initial condition $n(0) = 1$.

It is well known that under certain conditions the inequality $\sum_n p_n(t) < 1$ holds. The physical interpretation of this inequality is that there is a positive probability that an infinite number of events occur in finite time t.

We consider here the case where $\lambda_0 = 0$; if $\mu_1 > 0$ the state $n = 0$ is an attainable absorbing barrier. A necessary and sufficient condition for the occurrence of the phenomenon in this case is that the series

$$\sum_m \left(\frac{1}{\lambda_m} + \frac{\mu_m}{\lambda_m \lambda_{m-1}} + \cdots + \frac{\mu_m \cdots \mu_2}{\lambda_m \lambda_{m-1} \cdots \lambda_1} \right) \tag{1.1}$$

shall converge.

This result has been obtained in various equivalent forms by D. G. Kendall (unpublished, quoted by Bartlett [1]), Dobrusin [3], Karlin and McGregor [5], and Reuter and Ledermann [6].

This paper will present a simpler derivation of the result, which will at the same time emphasize the physical significance of the terms of the series.

2. Passage Times. We shall denote by τ_m the time taken for n to increase from m to $m + 1$, and consider the expected time $\bar{\tau}_m$ of such a change. If $\mu_1 > 0$ it is necessary to interpret the $\bar{\tau}_m$ as conditional expected times, conditional upon non-absorption.

THEOREM 1. *$\bar{\tau}_m$ is given by the recursion formula*

$$\bar{\tau}_m = \frac{1}{\lambda_m} + \frac{\mu_m}{\lambda_m} \bar{\tau}_{m-1}. \tag{2.1}$$

PROOF. The probability density function for the time t elapsing until the occurrence of the first event after the population size has reached m is

$$f(t) = (\lambda_m + \mu_m) \exp [-(\lambda_m + \mu_m)t]. \tag{2.2}$$

The expected value of t is thus $1/(\lambda_m + \mu_m)$. Such an event has probability $\lambda_m/(\lambda_m + \mu_m)$ of being a birth, in which case the population has passed from m to $m + 1$ as required, and probability $\mu_m/(\lambda_m + \mu_m)$ of being a death, when the desired increase requires further passage from $m - 1$ to m and then from m to $m + 1$.

We thus have

$$\bar{\tau}_m = \frac{\lambda_m}{\lambda_m + \mu_m} \frac{1}{\lambda_m + \mu_m} + \frac{\mu_m}{\lambda_m + \mu_m} \left(\frac{1}{\lambda_m + \mu_m} + \bar{\tau}_{m-1} + \bar{\tau}_m \right), \tag{2.3}$$

whence

$$\bar{\tau}_m = \frac{1}{\lambda_m} + \frac{\mu_m}{\lambda_m} \tau_{m-1}. \tag{2.4}$$

It follows that

$$\bar{\tau}_1 = \frac{1}{\lambda_1}, \qquad \bar{\tau}_2 = \frac{1}{\lambda_2} + \frac{\mu_2}{\lambda_2\lambda_1}, \cdots, \tag{2.5}$$

$$\bar{\tau}_m = \frac{1}{\lambda_m} + \frac{\mu_m}{\lambda_m\lambda_{m-1}} + \cdots + \frac{\mu_m \cdots \mu_2}{\lambda_m \cdots \lambda_1}. \tag{2.6}$$

If t_∞ denotes the time of passage to infinity, its expected value is given by

$$\bar{t}_\infty = \sum_n \bar{\tau}_n. \tag{2.7}$$

3. Divergence of the Process. We proceed to obtain the main results.

THEOREM 2. *If $\bar{t}_\infty$ is finite, there are values of t for which $\sum_n p_n(t) < 1$.*

PROOF. $\sum_n p_n(t) = 1$ implies that the probability that $t_\infty < t$ is zero, which in turn implies that

$$P(t_\infty > t) = 1. \tag{3.1}$$

Using Cramér's generalization of the Tchebycheff inequality [1], we have for all t,

$$P(t_\infty \geqq t) \leqq \frac{E(t_\infty)}{t} = \frac{\bar{t}_\infty}{t}, \tag{3.2}$$

so that for $t > \bar{t}_\infty$

$$\sum_{n=0}^{\infty} p_n(t) = P(t_\infty \geqq t) \leqq \frac{\bar{t}_\infty}{t} < 1, \tag{3.3}$$

and indeed, by taking t large enough, $\sum_{n=0}^{\infty} p_n(t)$ may be made as small as we wish. Thus, if $\bar{t}_\infty$ is finite, then for all $t > \bar{t}_\infty$, $\sum_{n=0}^{\infty} p_n(t) < 1$.

THEOREM 3. *If there is a finite time τ such that $\sum_n p_n(\tau) < 1$, then $\bar{t}_\infty$ is finite.*

PROOF. Suppose that

$$p_{1\infty}(\tau) = 1 - \sum_0^{\infty} p_n(\tau) = \alpha > 0; \tag{3.4}$$

then

$$P[n(\tau) < \infty] = 1 - \alpha \quad \text{and} \quad p_{i\infty}(\tau) \geqq \alpha, \qquad i \geqq 1, \tag{3.5}$$

$$P[n(m\tau) < \infty] \leqq (1 - \alpha)^m, \tag{3.6}$$

so that

$$P[n(m\tau) < \infty, n((m+1)\tau) = \infty] \leqq (1 - \alpha)^m; \tag{3.7}$$

thus

$$\begin{aligned} \bar{t}_\infty &\leqq \sum_{m=0}^{\infty} (m+1)\tau P[n(m\tau) < \infty, n((m+1)\tau) = \infty] \\ &\leqq \sum_{m=0}^{\infty} (m+1)\tau(1-\alpha)^m = \tau \sum_{m=0}^{\infty} (m+1)(1-\alpha)^m. \end{aligned} \tag{3.8}$$

But the series $\sum (m+1)x^m$ converges for $|x| < 1$, therefore $\bar{t}_\infty$ is finite.

COROLLARY 3.1. *A necessary and sufficient condition for the process to be divergent is that $\bar{t}_\infty$ shall be finite.*

The result of (1.1) follows immediately.

COROLLARY 3.2. *For a birth and death process with no lower absorbing barrier $P(t_\infty < \infty)$ is either zero or 1.*

PROOF. If $\bar{t}_\infty$ is finite then, from Theorem 2, we have for all $t > \bar{t}_\infty$

$$P(t_\infty > t) \leqq \frac{\bar{t}_\infty}{t}$$

But $(\bar{t}_\infty/t) \to 0$ as $t \to \infty$ so that

$$\lim_{t\to\infty} P(t_\infty < t) = 1, \quad \text{or equivalently} \quad \lim_{t\to\infty} \sum_n p_n(t) = 0. \tag{3.9}$$

It follows immediately from Theorem 3 that, if $P(t_\infty < \infty)$ is not zero, then $\bar{t}_\infty$ is finite, so that the probability must be 1.

4. Acknowledgements. I wish to thank Professor Casper Goffman for his assistance and advice during the direction of this work; I wish also to thank the referees for their helpful suggestions and for drawing my attention to references [1], [3] and [5].

REFERENCES

[1] M. S. BARTLETT, *Stochastic Processes*, Cambridge University Press, 1955, p. 88.
[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 182.
[3] R. L. DOBRUSIN, *Uspehi Matem. Nauk* (N.S.) 7 (1952) No. 6 (52) (185–191). Abstract in *Math. Rev.*, Vol. 14, p. 567.
[4] W. FELLER, *Probability Theory and Its Applications*, John Wiley & Sons, New York, 1950, pp. 371–373.
[5] S. KARLIN AND J. MCGREGOR, "Representation of a class of stochastic processes," *Proc. Nat. Acad. Sci.*, Vol. 41, pp. 387–391.
[6] G. E. H. REUTER AND W. LEDERMANN, "On the differential equations for the transition probabilities of Markov processes with enumerably many states," *Proc. Cambridge Philos. Soc.*, Vol. 49 (1953), pp. 247–262.

A REGRESSION ANALYSIS USING THE INVARIANCE METHOD

BY D. A. S. FRASER

University of Toronto

1. Summary. The invariance method is applied to a regression problem for which the "errors" have a rectangular distribution. The invariance method can also be applied to produce good estimates for the regression problem when the "errors" form a sample from any fixed distribution.

Received November 4, 1955; revised November 26, 1956.

2. Introduction. The invariance method is discussed in, for example, Blackwell and Girshick [1]. We summarize briefly its form for estimation. Let θ be a parameter that indexes the probability distributions and let there be a group of transformations s on the sample space that leaves the *class* of probability distributions unchanged. Suppose that the group of transformations is such that any of the probability distributions can be transformed into any other. This implies that the risk function for any invariant procedure is constant valued. Let $m(x)$ label the invariant subsets on the sample space for x. If s^* is the transformation on the parameter space corresponding to the transformation on the sample space, then it is easily seen that minimum risk estimator $f(x)$ may be found from any invariant estimator $f_0(x)$ by finding a transformation s_m^* for each m such that

$$E_{\theta_0}\{W(s_m^* f_0(x), \theta_0) \mid m(x) = m\}$$

is minimized for any fixed θ_0 where $W(f, \theta)$ is an invariant loss function and $f(x) = s_{m(x)}^* f_0(x)$.

3. A regression problem. This problem was suggested by Prof. E. G. Olds. Let $Y_1, \cdots, Y_n$ be real valued random variables with the following structure:

$$Y_i = \sum_1^r \beta_j x_{ij} + U_i, \tag{1}$$

where $U_1, \cdots, U_n$ are independent random variables and each is uniformly distributed with mean 0 and known range δ. In vector notation we have $\mathbf{Y} = \sum \beta_j \mathbf{x}_j + \mathbf{U}$. The x_{ij} are given numbers and the β_j are known regression coefficients. The problem is to obtain good estimates of the regression coefficients. In this section we find invariant estimators with minimum variances.

To simplify the notation we consider the case having $r = 2$ and

$$Y_i = \alpha + \beta x_i + U_i, \tag{2}$$

where the x_i have been adjusted so that $\sum x_i = 0$. Let the loss function for an estimate (f, g) of (α, β) be a weighted sum of the squared errors:

$$W(f, g; \alpha, \beta) = p(f - \alpha)^2 + q(g - \beta)^2 \qquad 0 \leqq p, q. \tag{3}$$

As a class of transformations consider

$$\{y_i' = y_i + a + bx_i (i = 1, \cdots, n) \mid (a, b) \varepsilon R^2\}. \tag{4}$$

It is straightforward to see that this class of transformations is an invariant class. Also the induced group of transformations on the parameter space is easily seen to be

$$\{\alpha' = \alpha + a, \beta' = \beta + b \mid (a, b) \varepsilon R^2\}. \tag{5}$$

Any $(\alpha, \beta) \varepsilon R^2$ can be transformed into any other point $(\alpha', \beta') \varepsilon R^2$; hence the group leaves no set invariant. Now, restricting ourselves to invariant estimators, we find that an estimator $(f(\mathbf{y}), g(\mathbf{y}))$ for (α, β) has the value

$$[f(\mathbf{y}^0) + a, g(\mathbf{y}_0) + b] \tag{6}$$

at the point

$$(y_1^0, \cdots, y_n^0) + a(1, \cdots, 1) + b(x_1, \cdots, x_n).$$

For convenience in describing (6) we introduce new coordinates in R^n, say $w_1, \cdots, w_n$, using $(1, \cdots, 1)$ and $(x_1, \cdots, x_n)$ as the unit points for the first two coordinates w_1, w_2 and any $(n-2)$ orthogonal vectors for the remaining coordinates. Then letting f^*, g^* be the functions f, g expressed in terms of the new coordinates, we obtain from (6):

$$\begin{aligned} f^*(a, b, w_3, \cdots, w_n) &= a + f^*(0, 0, w_3, \cdots, w_n), \\ g^*(a, b, w_3, \cdots, w_n) &= b + g^*(0, 0, w_3, \cdots, w_n). \end{aligned} \tag{7}$$

We need minimize the risk only for a single parameter value, say $\alpha = 0, \beta = 0$, and it will be uniformly minimized. The risk when $\alpha = 0, \beta = 0$ is

$$\begin{aligned} k \int_C & [p(a + f^*(0, 0, w_3, \cdots, w_n))^2 \\ & + q(b + g^*(0, 0, w_3, \cdots, w_n))^2]\, da\, db\, dw_3, \cdots, dw_n \\ = kp \int_C & [a + f^*(0, 0, w_3, \cdots, w_n)]^2\, da\, db\, dw_3, \cdots, dw_n \\ & + kp \int_C [b + g^*(0, 0, w_3, \cdots, w_n)]^2\, da\, db\, dw_3, \cdots, dw_n, \end{aligned} \tag{8}$$

where k is the constant value of the Jacobian from $y_1, \cdots, y_n$ to $w, \cdots, w$ and C is the set of values of $(a, b, w_3, \cdots, w_n)$ corresponding to the "cube" $[-\delta/2, \delta/2]^n$ in the coordinates $y_1, \cdots, y_n$. It is easily seen that the values for $f^*(0, 0, w_3, \cdots, w_n) g^*(0, 0, w_3, \cdots, w_n)$ which minimize the risk are such that $[-f^*(0, 0, w_3, \cdots, w_n), -g^*(0, 0, w_3, \cdots, w_n)]$ is the center of gravity of the set

$$C(w_3, \cdots, w_n) = \{(a, b) \mid (a, b, w_3, \cdots, w_n)\, \varepsilon\, C\},$$

the $w_3, \cdots, w_n$ section of C. This choice of f^*, g^* produces the minimum risk invariant estimate for α, β.

The determination of the values of the functions $f^*(0, 0, w_3, \cdots, w_n)$, $g^*(0, 0, w_3, \cdots, w_n)$ can, however, be simplified. If we change the sign of the coordinates of all points in $C(w_3, \cdots, w_n)$, then the center of gravity of the new set will have as coordinates the minimizing values $f^*(0, 0, w_3, \cdots, w_n)$, $g^*(0, 0, w_3, \cdots, w_n)$, determined in the paragraph above. This altered set has, however, a simple interpretation. It is the set of points (a, b) such that the cube C, shifted to have center at $(a, b, 0, \cdots, 0)$, contains the point $(0, 0, w_3, \cdots, w_n)$. Similarly, the value of the estimator,

$$[a' + f^*(0, 0, w_3, \cdots, w_n), b' + g^*(0, 0, w, \cdots, w)],$$

or $[f^*(a', b', w_3, \cdots, w_n), g^*(a', b', w_3, \cdots, w_n)]$, is the center of gravity of the points (a, b) for which the cube C, shifted to have center at $(a, b, 0, \cdots, 0)$, contains the point $(a', b', w_3, \cdots, w_n)$. However, it is equivalent to state that the estimate of (α, β) is the center of gravity of the points (a, b) for which the line $y = a + bx$ is a possible regression line for the observed points $(x_1, y_1), \cdots, (x_n, y_n)$, i.e., for which $y = a + bx$ is within $\delta/2$ vertically of each point $(x_1, y_1), \cdots, (x_n, y_n)$.

It is of interest to note that the estimates of α and β are $\bar{y}$ and $\sum y_i x_i / \sum x_i^2$ plus corrections which depend only on the deviations from the usual regression line. This is essentially the invariance requirement.

The methods of Section 3 up to formulas (7) and (8) may be applied in much the same manner to any regression problem for which the errors are a sample from some given fixed distribution.

REFERENCE

[1] DAVID BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley & Sons, New York, 1954.

ON DISCRETE VARIABLES WHOSE SUM IS ABSOLUTELY CONTINUOUS[1]

BY DAVID BLACKWELL

University of California, Berkeley

1. Summary. If $\{Z_n\}$, $n = 1, 2, \ldots$ is a stationary stochastic process with D states $0, 1, \cdots, D - 1$, and $X = \sum_1^\infty Z_k/D^n$, Harris [1] has shown that the distribution of X is absolutely continuous if and only if the Z_n are independent and uniformly distributed over $0, 1, \cdots, D - 1$, i.e., if and only if the distribution of X is uniform on the unit interval. In this note we show that if $\{Z_n\}$, $n = 1, 2, \cdots$ is any stochastic process with D states $0, 1, \cdots, D - 1$ such that $X = \sum_1^\infty Z_n/D^n$ has an absolutely continuous distribution, then the conditional distribution of $R_k = \sum_{n=1}^\infty Z_{k+n}/D^n$ given $Z_1, \cdots, Z_k$ converges to the uniform distribution on the unit interval with probability 1 as $k \to \infty$. It follows that the unconditional distribution of R_k converges to the uniform distribution as $k \to \infty$. Since if $\{Z_n\}$ is stationary the distribution of R_k is independent of k, the result of Harris follows.

2. Proof of the theorem.

THEOREM. *If $\{Z_n\}$, $n = 1, 2, \cdots$ is a sequence of random variables, each assuming only values $0, 1, \cdots, D - 1$ such that $X = \sum_1^\infty Z_n/D^n$ has an absolutely continuous distribution, and*

$$0 < \lambda \leqq 1, \text{ then } U_k(\lambda) = P\left(\sum_1^\infty Z_{k+n}/D^n < \lambda \mid Z_1, \cdots, Z_k\right) \to \lambda$$

with probability 1 as $k \to \infty$.

Received July 18, 1956.

[1] This paper was prepared with the partial support of the Air Research and Development Command under contract AF 41(657)-29), with the USAF School of Aviation Medicine.

PROOF. Say X has density p with respect to Lebesgue measure on the unit interval. Then

$$U_k(\lambda) = \lambda\, d(y_k(X), \lambda D^{-k})/d(y_k(X), D^{-k}),$$

where $y_k(s) = mD^{-k}$ for $mD^{-k} \leqq s < (m+1)D^{-k}$, $m = 0, 1, \cdots, D^k - 1$, and $d(a, h) = h^{-1}\int_a^{a+h} p(s)\, ds$.

We must show that

$$\lambda d(y_k(s), \lambda D^{-k})/d(y_k(s), D^{-k}) \to \lambda$$

for almost all s (Lebesgue measure) for which $p(s) > 0$, and this will follow from

$$d(y_k(s), \lambda D^{-k}) \to p(s) \qquad \text{a.e.} \tag{1}$$

Now a basic theorem of real variable theory asserts that

$$d(s, h) \to p(s) \qquad \text{a.e.} \tag{2}$$

as $h \to 0$. Let $a_k(s) = (s - y_k(s))/\lambda D^{-k}$.

Then

$$\begin{aligned} d(y_k(s), \lambda D^{-k}) &= a_k(s)\, d(s, y_k(s) - s) + [1 - a_k(s)]\, d(s, y_k(s) + \lambda D^{-k} - s) \\ &= a_k(s)[d(s, y_k(s) - s) - d(s, y_k(s) + \lambda D^{-k} - s)] \\ &\qquad + d(s, y_k(s) + \lambda D^{-k} - s). \end{aligned} \tag{3}$$

Since $a_k(s)$ is bounded, letting $k \to \infty$ in (3) and using (2) yields (1), and the proof is complete.

REFERENCE

[1] T. E. HARRIS, "On chains of infinite order," *Pacific J. Math.*, Vol. 5 (1955), pp. 707–724.

A PROOF THAT THE SEQUENTIAL PROBABILITY RATIO TEST (S.P.R.T.) OF THE GENERAL LINEAR HYPOTHESIS TERMINATES WITH PROBABILITY UNITY

BY W. D. RAY

British Coal Utilisation Research Association

1. Introduction. It can be shown [1] [2] that the S.P.R.T. of the general linear hypothesis resolves itself into the following form of procedure: Continue sampling at stage (n) if

$$\frac{\beta}{1-\alpha} < e^{-\lambda^{(n)}/2} M\left(\alpha(n), \gamma; \frac{\frac{1}{2}\lambda^{(n)} G^{(n)}}{1 + G^{(n)}}\right) < \frac{1-\beta}{\alpha} \cdots; \tag{1}$$

otherwise accept or reject the null hypothesis depending upon whether the left-hand or right-hand inequality is violated.

Received June 14, 1956; revised September 4, 1956.

$\lambda^{(n)}$ characterizes the alternative hypothesis, $\alpha(n)$ is half the sum of the degrees of freedom of the numerator and denominator of the test criterion $G^{(n)} = S_b/S_a$, and γ is half the degrees of freedom of S_b.

α, β are the probabilities of error of the first and second kind respectively.

$\lambda^{(n)}$, $\alpha(n)$ are each linear functions of n, the number of observations taken, γ is a fixed positive constant (where $\alpha(n) > \gamma > 0$), and

$$M(\alpha, \gamma; u) = \sum_{r=0}^{\infty} \frac{\Gamma(\gamma)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + r)}{\Gamma(\gamma + r)} \frac{u^r}{r!}.$$

Sampling is terminated whenever $G^{(n)} \leqq \underline{G}^{(n)}$ or $\geqq \bar{G}^{(n)}$, where $\underline{G}^{(n)}$, $\bar{G}^{(n)}$ are solutions of the equations

$$(2) \qquad f_n(G^{(n)}) = e^{-\lambda^{(n)}/2} M\left(\alpha(n), \gamma; \frac{\frac{1}{2}\lambda^{(n)} G^{(n)}}{1 + G^{(n)}}\right) = A, B \text{ respectively } \cdots,$$

where $A = \beta/(1 - \alpha), B = (1 - \beta)/\alpha$.

2. Proof that the test terminates with probability unity. It will be sufficient to prove that as $n \to \infty$, $\underline{G}^{(n)} \to \bar{G}^{(n)} \to G_0$, say. Now

$$f_n'(G) = e^{-\lambda/2} \frac{\alpha}{\gamma} M\left(\alpha + 1, \gamma + 1; \frac{\frac{1}{2}\lambda G}{1 + G}\right) \frac{\frac{1}{2}\lambda}{(1 + G)^2}$$

and

$$(3) \qquad f_n(G) = e^{-\lambda/2} M\left(\alpha, \gamma; \frac{\frac{1}{2}\lambda G}{1 + G}\right) \cdots.$$

From a recurrence relation of the Confluent Hypergeometric Function $M(\alpha, \gamma; u)$ it can be shown that

$$\frac{\gamma}{\alpha} < \frac{M(\alpha + 1, \gamma + 1; u)}{M(\alpha, \gamma; u)} < 1 \qquad (\text{for } u > 0),$$

from which it follows that

$$(4) \qquad \frac{\frac{1}{2}\lambda f_n(G)}{(1 + G)^2} < f_n'(G) < \frac{(\frac{1}{2}\lambda\alpha/\gamma) f_n(G)}{(1 + G)^2} \qquad \text{for all } G > 0 \cdots.$$

Let $g_n(G) = \log_e f_n(G)$. Then from (4) it follows that for $G > 0$,

$$\frac{\frac{1}{2}\lambda}{(1 + G)^2} < g_n'(G) < \frac{\frac{1}{2}\lambda}{(1 + G)^2} \frac{\alpha}{\gamma}.$$

Since $\lambda \to \infty$ as $n \to \infty$, this inequality shows that $g_n'(G) \to \infty$ as $n \to \infty$. Further, since $g_n(G)$ is a positive strictly increasing continuous function of G, it follows that there can exist at most one value of G, say G_0, where $g_n(G)$ does not become infinite as $n \to \infty$. Consequently $g_n(G) \to -\infty$ for $G < G_0$ and $g_n(G) \to +\infty$ for $G > G_0$. In terms of $f_n(G)$, this implies that $f_n(G) \to 0$ for $G < G_0$, and $f_n(G) \to \infty$ for $G > G_0$. This in turn implies that $G^{(n)} \to \underline{G}^{(n)} \to G_0$, and sampling must therefore terminate.

If there does not exist a finite G_0 for which $g_n(G)$ does not become infinite, then $g_n(G)$ becomes infinite for all $G > 0$. Thus $g_n(G)$ either becomes infinite for all $G > 0$ or approaches zero for all $G > 0$. In the first case, sampling will terminate because $f_n(G) > B$ for sufficiently large n for all $G > 0$; and in the second case too, since $f_n(G) < A$ for sufficiently large n for all $G > 0$.

3. Comments. It has been possible to obtain an upper bound for the limiting value G_0 but not to obtain its value uniquely. David and Kruskal [3] have provided a solution to the same problem for the sequential t-test.

4. Acknowledgement. I am most grateful to Dr. N. L. Johnson for his guidance during research on this problem, to the referee for his comments, and to the British Coal Utilisation Research Association for permission to publish this paper.

REFERENCES

[1] N. L. JOHNSON, "Some notes on the application of sequential methods in the analysis of variance," *Ann. Math. Stat.*, Vol. 24, (1953), pp. 614–623.

[2] P. G. HOEL, "On a sequential test for the general linear hypothesis," *Ann. Math. Stat.*, Vol. 26, (1955), pp. 136–139.

[3] H. T. DAVID AND W. H. KRUSKAL, "The WAGR Sequential t-Test reaches a decision with probability one," *Ann. Math. Stat.*, Vol. 27, (1956), pp. 797–805.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Washington meeting of the Institute, March 7–9, 1957)

1. **Synchronization of Trajectory Images of Ballistic Missiles and the Timing Record of the Ground Telemetry Recording System,** HARRY P. HARTKEMEIER, Stanford University, (introduced by Paul R. Rider).

In order to compute the position, velocity, and acceleration of a missile, it is necessary to synchronize the image pattern from ballistic camera plate records and the timing record of the ground telemetry recording system. In the past this has been done by personal inspection. This takes too much time; consequently, a method by which the two records may be matched by high-speed electronic computers is required to speed up the work.

The missile is equipped with two strobe lights, one on each side, which are supposed to flash simultaneously when scheduled to do so by a programmer. Inside the missile there is a timing generator controlled by a tape punched according to a coding pattern. When the timing generator sends a signal for the strobe lights to flash, it also sends a signal simultaneously to the telemetry transmitter. This signal reaches the ground recording telemetry system through a radio link. A method of matching these two records by using correlation technique and an electronic computer is presented. (Received November 6, 1956.)

2. **Maximum Likelihood Estimates in a Simple Queue,** A. BRUCE CLARKE, University of Michigan, (By Title).

A simple stationary queueing process is a process having a Poisson input (with parameter λ), and a negative exponential service time (with mean $1/\mu$, $\mu > \lambda$). Let ν = the initial

queue size, x_i = the time of the ith arrival, y_i = the "busy time" up to the ith departure. The sequences $\{x_i\}$ and $\{y_i\}$ then represent the transition times of *independent* Poisson processes (parameters λ and μ), and $\{\nu, \{x_i\}, \{y_i\}$ together characterize the process. By observing the process for a fixed "busy time" τ and using the above comment, maximum likelihood estimates for λ and μ may be obtained in terms of ν, m = the total number of departures, T = the time of the mth departure, and n = the total number of arrivals up to time T. Under certain conditions these estimates of λ and μ may be approximated by $(n + \nu)/T$ and $(m - \nu)\tau$. (Received November 12, 1956.)

3. A Rank Order Test for Trend in Correlated Means, Ardie Lubin, Walter Reed Army Institute of Research, Washington, D. C.

In many experiments the major interest is not in the amount of difference caused by the treatments but the rank-order which results. This is especially true when successive measurements are made on the same subject, and the "treatments" are simply varying amounts of fatigue, sleep loss, etc., i.e., some function of time. For such studies the null hypothesis is that no trend exists and generally the only alternative hypothesis is a rank-order that can be specified by the experimenter.

A. R. Jonckheere ("A distribution-free k-sample test against ordered alternatives," *Biometrika*, Vol. 41 (1954)) has used Kendall's tau to obtain a general statistic, P, for testing the agreement between a hypothesized rank-order for n objects or scores and a set of observed rankings of the n scores by m judges. From this general approach, he derives a test for trend as a special case.

As an alternative to Jonckheere's P, a statistic J based on Spearman's $S(d^2)$ is examined. It is the sum of the $S(d^2)$ values computed between the m observed rankings of the n scores and the hypothesized rank-order of the n scores. K, the average rank order correlation between the m rankings and the hypothesized rank-order, is a simple algebraic function of J.

It is shown that J is slightly more sensitive than Jonckheere's P statistic for small values of n, but that P tends to normality faster than J. (Received November 13, 1956.)

4. On the Stochastic Structure of Minkowski-Leontief Systems, David Rosenblatt, American University.

A linear system $x(I - A) = w$ is said to be of *Minkowski-Leontief type* if A is a finite nonnegative square matrix of order n with no row sum exceeding unity and x, w are nonnegative row vectors. A non-null solution x of such a system is called *admissible*. Theorem: *Every system of Minkowski-Leontief type $x(I - A) = w$ which exhibits at least one admissible solution is equivalent to a unique system $\bar{x}(I - \bar{A}) = \theta$, where $\bar{A}$ is a stochastic matrix depending on A and w and θ is a null vector of dimension at most $n + 1$. Every admissible solution of $x(I - A) = w$ (appropriately extended or contracted) is proportional to a convex linear combination of the stationary stochastic vectors of $\bar{A}$.* If A is nonstochastic, $w \neq \theta_n$, let $\bar{A}$ denote the matrix $\begin{Vmatrix} A & b \\ w^* & 0 \end{Vmatrix}$ where $b = (I - A)e'$, $w^* = \lambda_w^{-1} w$, $\lambda_w = we'$, and e is the row vector with all elements unity. If $(I - A)^{-1}$ exists and $w \neq \theta_n$, there exists a single ergodic set of indices; if w is positive the stationary vector of $\bar{A}$ is positive. Clearly,

$$\bar{x} = (w(I - A)^{-1}, \lambda_w)$$

If $w \neq \theta_n$ and $(I - A)$ is singular, $\bar{A}$ is taken as $\begin{Vmatrix} A_r & 0 \\ 0 & I_{n-r} \end{Vmatrix}$ where A_r is the largest stochastic principal submatrix of A. Systems of the present type occur in economic input-output analysis and generally in socio-physical models based on "balanced-margin" tables, i.e., nonnegative square matrices X such that $eX = eX'$. (Received November 20, 1956.)

5. **The Joint Distribution of a Set of Sufficient Statistics for the Parameters of a Simple Telephone Exchange Model,** VÁCLAV EDVARD BENEŠ, Bell Telephone Laboratories, (By Title).

This paper considers a simple telephone exchange model which has an infinite number of trunks and in which the traffic depends on two parameters, the calling-rate and the mean holding-time. It is desired to estimate these parameters by observing the model continuously during a finite interval, and noting the calling-time and hang-up time of each call, insofar as these times fall within the interval. It is shown that the resulting information may, for the purpose of this estimate, be reduced without loss to four statistics. These statistics are the number of calls found at the start of observation, the number of calls arriving during observation, the number of calls leaving during observation, and the average number of calls existing during the interval of observation. The joint distribution of these sufficient statistics is determined (in principle) by deriving a generating function for it. From this generating function, the means, variances, covariances, and correlation coefficients are obtained. Various estimators for the parameters of the model are compared, and some of their distributions, means, and variances presented. (Received November 29, 1956.)

6. **On the Stochastic Structure of Minkowski-Leontief Systems, II,** DAVID ROSENBLATT, American University, (By Title).

Consider a system $x(I - A) = w$ of Minkowski-Leontief type such that $(I - A)^{-1}$ exists. Clearly, $(I - A)^{-1}$ exists if and only if A contains no stochastic principal submatrix. In a static economic input-output context the element a_{ij} is designated as the input (per unit output) to industry or activity i procured from industry j; w_j, x_j are respectively final output and total output (or activity level) of the jth industry. Consider the uniquely corresponding system $\bar{x}(I - \bar{A}) = \theta$, where $\bar{A}$ is stochastic. The unique stationary stochastic vector of $\bar{A}$ is given by $(p_{n+1}w^*(I - A)^{-1}, p_{n+1})$. The "multiplier" $\mu = \sum_{j=1}^{n+1} x_j/\lambda$ is given by $1/p_{n+1}$, where $\lambda = x_{n+1} = wc'$. Given a nonsingular matrix $(I - A)$, the following relation holds in components of an admissible solution for any w: $\sum_{j=1}^{n} (1 - r_j)x_j - x_{n+1} = 0$, where $x_{n+1} = wc'$ and r_j is the jth row sum in A. The latter relation is the technical production-possibility function of the economy in an input-output sense; $-\Delta x_j/\Delta x_k = (1 - r_k)/(1 - r_j)$, $\Delta x_j/\Delta x_{n+1} = 1/(1 - r_j)$, $j \neq k, j, k = 1, \ldots, n$, are the invariant "substitution ratios" of the system, obviously independent of w. Let x be an admissible solution of $x(I - A) = w$, $(I - A)$ singular or not, and let $D(x, \lambda)$ be a diagonal matrix with components of x and λ on the diagonal. Then $D(x, \lambda)\bar{A}$ is a "balanced-margin" table. Consistent with a noted "substitution" result, $\sum_{j=1}^{n} K_j w_j = x_{n+1} = \lambda$, where $K_j = 1$ for all j independently of w. (Received December 17, 1956.)

7. **A Further Contribution to the Theory of Systematic Statistics,** JUNJIRO OGAWA, University of North Carolina.

Up to 1945 the main interest of statistical estimation has been in the "efficient estimator," but from the point of view of practical use, it seems reasonable to inquire whether comparable results could have been obtained by a smaller expenditure. F. Mosteller (1946) proposed the use of systematic statistics in this connection. The author (1952) developed a systematic theory of estimation and testing hypotheses with respect to the location and scale parameter of a population whose density depends on only these two parameters.

There are many cases in which the samples are by their very nature ordered in magnitude, for example in a life test of electric lamps. In such cases the population probabilty distributions are usually supposed to be exponential. Thus, at least for the exponential

distribution, estimation and testing of a hypothesis based upon systematic statistics are of great importance from the standpoint of practical application.

There will be presented in this paper the table of the optimum spacings of the selected sample quantiles, corresponding best estimators, and a discussion on the testing procedure of a statistical hypothesis on the scale parameter σ of the exponential distribution $f(x) = (1/\sigma)e^{-(x/\sigma)}$ for $x > 0$. (Received January 7, 1957.)

8. On the Stochastic Structure of Minkowski-Leontief Systems, III, David Rosenblatt, American University, (By Title).

Consider any system $x(I - A) = w$ of M-L type. The following "aggregation" problem is of interest. Let an *aggregation matrix* C be an $(n \times r)$ stochastic matrix of incidence type, $1 \leqq r < n$. Let $B = f(A)$ be a M-L matrix of order r. We consider conditions under which $\hat{x}AC = \hat{x}CB$ obtains for admissible solutions $\hat{x}$ of a system $x(I - A) = w$. The following case is of special interest. Let a *weight matrix* E be an $(n \times n)$ diagonal matrix with nonnegative entries on the principal diagonal. A *consolidation* of a matrix A of M-L type is an $(r \times r)$ matrix $B = B(A; C, E) = (C'EC)^{-1}C'EAC$, $1 \leqq r < n$. "Faithful consolidation" of a stochastic system $x(I - A) = 0$ is characterized from the standpoint of ergodic structure; the condition $AC = CB(A; C, E)$ is of particular interest. A general consolidation condition for M-L systems is related to the "combining-of-classes" condition of stochastic learning theory. The following is of economic interest: the existence of $(I - B)^{-1}$ does not in general imply the existence of $(I - A)^{-1}$, and conversely. In the static input-output model of II, the ergodic structure of $\bar{A}$ of the equivalent system (and the role of mean recurrence time $1/(p_{n+1})$) suggest that the stationary stochastic vector $\bar{y}$ of $\bar{A}$ be computed iteratively using successive powers of $\bar{A}$, yielding $\bar{x}$, in lieu of matrix inversion with or without consolidation; in most applications, $\lim_{k\to\infty}\bar{A}^k$ exists. (Received January 14, 1957.)

9. The Use of Incomplete Block Designs for Asymmetrical Factorial Arrangements, Marvin Zelen, National Bureau of Standards.

Let A_s $(s = 1, 2, \cdots, m)$ denote the sth factor in a m-factor factorial experiment such that A_s has m_s levels. Let $i = (i_1, i_2, \cdots, i_m)$ represent a particular experimental combination of the m-factors and let the mathematical model underlying the measurements be

$$y_{ij} = \mu + \sum_{s=1}^{m} (a_s)_{i_s} + \sum_{t=2}^{m}\sum_{s=1}^{t} (a_{st})_{i_s i_t} + \cdots + (a_{12\cdots m})_{i_{12\cdots m}} + b_j + \epsilon_{ij},$$

where $(a_s)_{i_s}$, $(a_{st})_{i_s i_t}$, $\cdots$, $(a_{12\cdots m})_{i_{12\cdots m}}$ represent the various main effects and interactions, b_j represents the block effect, and the ϵ_{ij} are NID $(0, \sigma^2)$. Algorithms are given for using the balanced incomplete and the group divisible designs for asymmetrical factorial arrangements. Let $M(s)$ be the square matrix (of dimension M_s) $M(s) = m_s I - J$ where J is a matrix having all elements unity, and define the direct product of p such matrices by $M(1, 2, \cdots, p) = [M(1) \times M(2) \times \cdots \times M(p)]$ $(p \leqq m)$. Then the variance-covariance matrix of a p-factor interaction for the G.D. case can be written as $M(1, 2, \cdots, p)\,\sigma^2/(E_t r v)$ $(t = 1$ or $2)$. For the BIBD, the same expression holds with $E_1 = E_2$. The correlations between the different interactions are all zero and since $M^2(1, 2, \cdots, p) = M(1, 2, \cdots, p)\prod_1^p m_s$, $[E_t r/\prod_1^p m_s]\sum (a_{12\cdots p})^2_{i_{12\cdots p}}$ follows a $\sigma^2\chi^2$ with $\prod_1^p (m_s - 1)$ degrees of freedom under the hypothesis of no p-factor interaction effects. (Received January 16, 1957.)

10. An Extension of the Cramér-Rao Inequality, John J. Gart, Virginia Polytechnic Institute, (By Title).

Consider a frequency function $f(x \mid \theta)$ where $\theta = (\theta_1, \theta_2, \cdots, \theta_s)$, the function being specified when θ is specified. The parameter θ has a density $g(\theta)$ independent of x. Let $\mathbf{X} =$

$(x_1, x_2, \cdots, x_n)$ be a random sample from a randomly chosen population having the specified frequency function. Then if $\phi = \prod_{i=1}^{n} f(x_i \mid \theta)$ and t_k (independent of θ) is an estimate of θ_k, $1 \leqq k \leqq s$, there follows a form similar to the Cramér-Rao Inequality, $EE[(t_k - \theta_k)^2 \mid \theta] \geqq \{E[E(t_k \mid \theta) - \theta_k]\}^2 + E^2\{[\partial E(t_k \mid \theta)]/\partial\theta_k\}\{EE[(\partial \ln\phi/\partial\theta_k)^2 \mid \theta]\}^{-1}$. The equality is reached if and only if t_k is an unbiased sufficient statistic having the normal distribution with constant variance. In this case the equality holds regardless of the form of $g(\theta)$. (Received January 17, 1957.)

11. Multivariate Analysis of Variance, S. N. Roy, University of North Carolina.

Consider a model under which we have stochastic variates $X(p \times n) = [x_1 \ldots x_n]p$ such that x_i's (for $i = 1, 2, \ldots, n$) are independent $N[E(x_i), \Sigma]$, $E(X') = A(n \times m) \times \xi(m \times p)$, A (to be called the design matrix) is a matrix of constants given by the design of the experiment, ξ is a matrix of unknown parameters, rank $(A) = r \leqq m < n$, $p \leqq n - r$ and Σ is an unknown dispersion matrix. Under this model suppose we have a testable hypothesis (the meaning and mathematical criterion for testability being discussed in the paper) $H_0: C(s \times m)\xi(m \times p)M(p \times q) = 0$ $(s \times q)$, where C and M (to be called the hypothesis matrices) are given such that rank $(C) = s \leqq r$ and rank $(M) = q \leqq p$. The alternative is $H: C\xi M = \eta$ $(s \times q)$ $(\neq 0)$. The test is that at a level α we accept H_0 if $c_{\max}(S^*S^{-1}) \leqq c_\alpha$ and reject H_0 otherwise, where S^* and S are matrices given (in the paper) in terms of X, A, C and M, $c_{\max}(T)$ denotes the largest root of a matrix with real nonnegative roots, and c_α is a constant depending on α, min (s, q) and $n - r$, which we can pick up from tables now under construction and expected to be published shortly. (Received January 17, 1957.)

12. Confidence Bounds Associated with Multivariate Analysis of Variance, S. N. Roy and R. Gnanadeshikan, University of North Carolina.

We start from the same set up as in *the previous paper*. The S^* and S (to be called respectively the dispersion matrix "due to the hypothesis" and the dispersion matrix "due to the error") are the exact analogs of the variance "due to the hypothesis" and that "due to the error" in the customary univariate analysis of variance. Given any level α, we can pick up a constant c_α from the tables mentioned in *the previous paper* and make, with a probability greater than or equal to $1 - \alpha$, the confidence interval statement: $c_{\max}^{1/2}(sS^*) - [sc_\alpha]^{1/2} \times c_{\max}^{1/2}(S) \leqq c_{\max}^{1/2}[\eta' U \eta] \leqq c_{\max}^{1/2}(sS^*) + [sc_\alpha]^{1/2} c_{\max}^{1/2}(S)$, where $U(s \times s)$ is a nonsingular matrix given (in the paper) in terms of A and C, and $c_{\max}^{1/2}[\eta' U \eta]$ is zero if and only if $\eta = 0$, i.e., H_0 is true. With a joint probability greater than or equal to $1 - \alpha$ we can also make simultaneous confidence interval statements including the one given above and others exactly similar to this but in terms of $S^{(i)}$, $S^{(i)^*}$, $\eta^{(i)}$ (for $i = 1, 2, \ldots, p$) and next in terms of $S^{(i,j)}$, $S^{(i,j)^*}$, $\eta^{(i,j)}$ (for $i \neq j = 1, 2, \ldots, p$), and so on, where $S^{(i)}$ and $S^{(i)^*}$ stand respectively for truncated matrices after cutting out the ith row and ith column from S and S^*, $\eta^{(i)}$ for η with the ith column cut out, $S^{(i,j)}$, $S^{(i,j)^*}$ for S and S^* with the ith and jth rows and columns cut out, $\eta^{(i,j)}$ for η with the ith and jth columns cut out, and so on. (Received January 17, 1957.)

13. Extension of Some Results Given by Mitra on "Statistical Analysis of Categorical Data," Earl Diamond, University of North Carolina.

This is a follow-up of two previous paper ([1] "Some non-parametric generalizations of analysis of variance and multivariate analysis" by S. N. Roy and S. K. Mitra, *Biometrika*, December, 1956, and [2] "Contributions to the statistical analysis of categorical data" by S. K. Mitra, North Carolina Institute of Statistics Mimeograph Series No. 142). We start from a product of multinomial distributions of the form $\phi = \prod_i [n_{0i}! \prod_j p_{ij}^{n_{ij}} / \prod_j n_{ij}!]$

with $\Sigma_i p_{ij} = 1$, $i = i_1 i_2 \dots i_k$; $j = j_1 j_2 \dots j_\ell$; $i_1 = 1, 2, \dots, r_1$; $\dots$; $i_k = 1, 2, \dots, r_k$; $j_1 \varepsilon (s_1)_{j_2 \cdots j_\ell}$ (a subset of s_1 depending on the subscript set $j_2 \dots j_\ell$); $j_2 \varepsilon (s_2)_{j_3 \cdots j_\ell}$; $\dots$; $j_{\ell-1} \varepsilon (s_{\ell-1})_{j_\ell}$ and $j_\ell = 1, 2, \dots, s_\ell$. We next consider two hypotheses $H_0^{(1)}: p_{ij} = f_{ij}^{(1)}(\theta_1, \dots, \theta_{t_1})$ subject to $g_m^{(1)}(\theta_1, \dots, \theta_{t_1}) = 0$ $(m = 1, 2, \dots, u_1 < t_1)$ and $H_0^{(2)}: p_{ij} = f_{ij}^{(2)}(\theta_1', \dots, \theta_{t_2}')$ subject to $g_m^{(2)}(\theta_1', \dots, \theta_{t_2}') = 0$ $(m = 1, 2, \dots, u_2 < t_2)$, where $t_1, t_2 <$ total number of cells $-$ total number of multinomial distributions. Each hypothesis is a composite one in which the θ's or θ''s are the nuisance parameters and $f_{ij}^{(1)}$, $g_m^{(1)}$, $f_{ij}^{(2)}$ and $g_m^{(2)}$ are known functions. Tests are taken over from Refs. [1] and [2], and the asymptotic powers of the tests and the conditions for asymptotic independence are derived which are extensions of similar conditions for more special cases discussed in [2]. (Received January 17, 1957.)

14. Testing of Hypotheses on a Mixture of Variates Some of Which are Continuous and the Rest Categorical, S. N. Roy and M. D. Moustafa, University of North Carolina.

We start from a $k + \ell$-variate distribution in which k variates are continuous and ℓ variates are categorical. The k variates are assumed to have a conditional multivariate normal distribution with respect to the ℓ categorical variates which are assumed to have a multinomial distribution. Appropriate hypotheses are framed in this situation, analogous to the customary hypotheses on a single multivariate normal distribution (or to those in Refs. [1] and [2] of the *previous abstract*), large sample tests of such hypotheses are developed and some of their properties studied. Next, instead of assuming a single multinomial distribution on the ℓ categorical variates, a product of multinomial distributions is assumed and hypotheses are framed in this situation analogous to the customary ones for several multivariate normal distributions or to those *in Refs. [1] and [2]*, and large sample tests of such hypotheses and some of their properties are studied. (Received January 17, 1957.)

15. On Statistics Independent of a Sufficient Statistic, Evan J. Williams, North Carolina State College.

It is shown that if, for a sample drawn from a population of values of x with distribution depending on a parameter θ, the statistic z is sufficient for θ, and g is any statistic whose distribution is independent of θ, then g and z are independently distributed. The method of proof is less sophisticated than that of Basu (*Sankhyā*, Vol. 15 (1955), p. 377).

The result has application to the normal distribution: the mean of a sample is distributed independently of any location-free statistic; and to the gamma distribution: the mean of a sample is distributed independently of any scale-free statistic. These well-known results follow since the sample mean is a sufficient statistic, in the former case for the location parameter, in the latter case for the scale parameter.

The limitations of the general result lie in the difficulty of deriving statistics independent of parameters other than location and scale parameters.

The connexion of the theorem with estimation theory is discussed. (Received January 17, 1957.)

16. Generalized Quantal Response in Biological Assay, John Gurland, Iowa State College.

The quantal (all-or-none) response in biological assay refers to a response in which one of two possible outcomes occurs. In a bioassay such as that of an insecticide based on mortality of the housefly, say, there are, however, three possible outcomes, namely, alive, moribund, dead. The present paper considers a generalized quantal response in which

two or more outcomes are possible. Whether one uses normits (cf. probits) or logits or other transformations, a general method of analyzing the data is developed which makes explicit use of all the possible outcomes and hence is more efficient than the common procedure of pooling some outcomes (for example, moribund and dead) in order to make the response all-or-none. Further, a technique analogous to that used in discriminant functions is suggested as a method which makes more efficient use of the data than the pooling method mentioned above. (Received January 21, 1957.)

17. **The Variance of Zero-Crossing Intervals,** J. A. McFadden, U. S. Naval Ordnance Laboratory, (introduced by Gilbert Lieberman).

Two expressions are given for the variance of the intervals between successive zeros of a random process. It is assumed that the successive intervals form a Markoff chain. If $x(t)$ is a random process, let $y(t) = 1$ when $x(t) \geqq 0$ and $y(t) = -1$ when $x(t) < 0$. Let β be the expected number of zeros per second and let κ be the correlation coefficient between two successive zero-crossing intervals. Then the variance is $\sigma^2 = (2A/\beta)(1 + \kappa)/(1 - \kappa)$, or alternatively, $\sigma^2 = [(1 + 2B)/\beta^2]\,(1 - \kappa)/(1 + \kappa)$, where $A = \int_0^\infty r(\tau)\,d\tau$ and $B = \int_0^\infty [Q(\tau) - \beta]\,d\tau$. $r(\tau)$ is the autocorrelation function of the process $y(t)$ and $Q(\tau)\,d\tau$ is the conditional probability of a zero between $t + \tau$ and $t + \tau + d\tau$, given a zero at time t. (Received January 21, 1957.)

18. **A Limit Theorem and Bounds for an Optional Stopping Probability,** Morris Skibinsky, Michigan State University, (By Title).

Let S_j be the standardized jth partial sum of a sequence of bounded independent, identically distributed random variables, K, a positive constant, and let

$$Q(m, n, K) = \Pr\{\max_{m \leqq j \leqq n} s_j \geqq K\}.$$

It is shown by elementary methods that if $\lim_{m\to\infty}[(n - K)/m^{1/2}] = 0$, then $\lim_{m\to\infty} Q(m, n, K) = 1 - \phi(K)$, where ϕ is the standard normal c.d.f. Certain steps in the proof are then used to obtain simple bounds for $Q(m, n, K)$ when the sequence of random variables is generated from Bernouli trials. (Received January 21, 1957.)

19. **A Limit Theorem of Cramér and Its Generalization,** Junjiro Ogawa, University of North Carolina, (By Title).

As a generalization of Doob's theorem, H. Cramér states the following theorem: *Suppose we have for every $\nu = 1, 2, \ldots$, $y_\nu = Ax_\nu + z_\nu$, where x_ν, y_ν and z_ν are n-dimensional random variables, while A is a matrix of order $(n \times n)$ with constant elements. Suppose further that as $\nu \to \infty$, the n-dimensional distribution of x_ν tends to a certain limiting distribution, while z_ν converges in probability to zero. Then y_ν has the limiting distribution defined by the linear transformation $y = Ax$, where x has the limiting distribution of the x_ν* (H. Cramér, *Mathematical Methods of Statistics*, Princeton, 1946, pp. 299–300). Cramér skips the proof of this theorem. In this paper, the complete proof of this theorem will be given and two theorems which are generalizations of this theorem and are useful in statistics will be proved. (Received January 22, 1957.)

20. **On the Mathematical Principles Underlying the Theory of the χ^2 Test,** Junjiro Ogawa, University of North Carolina, (By Title).

The rigorous proof of the theorem that the χ^2 statistic has the limiting chi-square distribution with degrees of freedom reduced by the number of the independent parameters

which were estimated from the frequency data, was first given by H. Cramér in his famous book *Mathematical Methods of Statistics*, Princeton (1946), but some steps of the proof were skipped. Later on S. N. Roy and S. K. Mitra (*Biometrika*, Vol. 43 (1956)) and S. K. Mitra (Thesis, University of North Carolina, 1956) reasoned along the same lines and got theorems adjusted to various physical situations. The purposes of this paper are to present a complete and self-contained proof of Cramér's theorem on the one hand, and on the other to explain how the proof of the related theorems got by S. N. Roy and S. K. Mitra could be thrown back on that of Cramér's theorem from the mathematical point of view. (Received January 22, 1957.)

21. **Minimization of Certain Integrals Subject to Linear Constraints, (Preliminary Report)**, C. H. KRAFT AND I. OLKIN, Michigan State University.

Let F be the class of measures f such that $E_f q_i(x) = a_i$, $i = 1, \ldots, n$ and $E_f H(f) < \infty$. The problem of minimizing $E_f H(f)$ over F has been treated by Shannon [*Bell System Technical Journal*, Vol. 27 (1948), pp. 623–656] for $H(f) = \log f$, $q(x) = x^2$ using calculus of variations, and by Weiss [*Ann. Math. Stat.*, Vol. 27 (1956), pp. 851–853] for $H(f) = f$, arbitrary square integrable $q_i(x)$.

The following considerations apply to these cases as well as others, e.g. $H(f) = f^p$. An inequality of the form $E_f H(f) \geqq T(f, g)$ for all densities g is available, where $T(g, g) = E_g H(g)$. $T(f, g)$ is constant for $f \in F$ if and only if $g(x) = \Sigma b_i q_i(x)$. The bound is attainable if the constants b_i can be chosen so that $g \in F$. These considerations extend the proofs to not necessarily dominated families F on any measure space. (Received January 23, 1957.)

22. **The Recovery of Intervariety Information,** BRADLEY BUCHER, Princeton University.

Assume, in the incomplete block model, $y_{ij} = m + b_i + v_j + e_{ij}$, that the block effects are independently distributed with mean 0 and variance β^2, the error terms e_{ij} are independently distributed with mean 0 and variance α^2, and that the variety effects $t_1, \ldots, t_k$, are fixed effects and that $t_{k+1}, \ldots, t_n$, are independently distributed with mean 0 and variance γ^2. Then in estimating any linear combination of the variety effects, say, $a_1 t_1 + a_2 t_2 + \ldots + a_k t_k$, we may make use of information among the varieties $t_{k+1}, \ldots, t_n$. Minimum variance linear unbiased estimates are obtained for such combinations for a large class of incomplete block designs. In general, these estimates have smaller variance than analogous estimates obtained using only inter- and intra-block recovery. For balanced incomplete blocks the estimate with intervariety recovery is shown to be the same as the combined intra- and inter-block estimate. Several techniques are developed which are useful for finding estimates using intervariety recovery. The problem of estimating γ^2 is discussed. Useful applications of the technique of intervariety recovery are considered. (Received January 24, 1957.)

23. **Some Uses of Quasi-Ranges II,** J. T. CHU AND F. C. LEONE; Case Institute of Technology AND C. W. TOPP, Fenn College, Cleveland, Ohio.

In "Some uses of quasi-ranges," (*Ann. Math. Stat.*, Vol. 28 (1957), No. 1), methods are given of using quasi-ranges to obtain confidence intervals for, and tests of hypotheses about, some measures of dispersion of a given distribution (such as the interquantile distance and the standard deviation). In this paper, further research is done on the selection of quasi-ranges for making inferences about the standard deviations of the normal, rectangular, and exponential distributions. The methods are also extended to the co-

efficient of variation, the difference and ratio of interquantile distances and standard deviations of two given distributions, etc. Tables are given to facilitate applications. Received January 24, 1957.)

24. **On Selecting a Subset Which Contains All Populations Better Than a Standard,** SHANTI S. GUPTA AND MILTON SOBEL, Bell Telephone Laboratories.

Populations $\pi_i (i = 0, 1, \ldots, p)$ are given with a common Koopman-Darmois distribution of known form differing only in the value of the unknown parameter $\tau_i (i = 1, 2, \ldots, p)$; cases of known and unknown (associated with the standard π_0) are treated separately. Location and scale parameter problems are both treated. In some problems π_i is defined as better than π_0 if $\tau_i > \tau_0$; in others if $\tau_i < \tau_0$. A procedure is given in each case for selecting a small subset so that, for any true configuration, the probability of including all π_i equal to or better than π_0 is at least P^*, $P^* < 1$ being preassigned. For the location parameter, with τ_0 unknown, the procedure is to retain all π_i with $\bar{w}_i = \sum_{j=1}^{n_i} w(x_{ij}) \geqq \bar{w}_0 - d/(n_i)^{1/2}$; here $\bar{w}_i$ is sufficient for $\tau_i (i = 0, 1, \ldots, p)$. For scale parameter problems, with smaller τ_i more preferable, the procedure retains all π_i with $\sum_{j=1}^{n_i} w(x_{ij}) \leqq (1 + d) \sum_{j=1}^{n_0} w(x_{0j})$. In several problems the value of d is computed and tables are given for different P^* and p-values; in others transformations are used to "normalize" the problem. The normal and chi-square distributions are used as applications. Problems involving binomial and Poisson distributions are treated separately with and without normalizing transformations. (Received January 24, 1957.)

25. **On the Relation Between Loss Functions and Significance Levels, (Preliminary Report),** H. ROBERT VAN DER VAART, North Carolina State College and Leiden University.

Consider a one-parameter family $\{P_\theta\}$ of probability distributions. Be it asked to test $H_0: \theta = \theta_0$, against $H_1: \theta > \theta_0$. Define a loss function $L = l_0$ if H_0 is rejected when true, $L = l_1$ if H_1 is rejected when true, $L = 0$ otherwise. Suppose a family $\{w_t\}$ of subsets of the sample space is given, $P_\theta(w_t)$ being a monotonous increasing function of θ for each w_t. Then selecting a critical region w_* such that $P_\theta(w_*)$ has some fixed value α is a classical procedure, known to be minimax relatively to L provided $\alpha = l_1/(l_0 + l_1)$ (Sverdrup, 1953; Ruist, 1954). However most statisticians, while fixing $P_\theta(w_*) = \alpha$ for $\theta = \theta_0$, really want(ed) to reject H_0 only if θ differs materially from θ_0, say if $\theta > \theta_1 > \theta_0$ (cf. also Hodges and Lehmann, 1954), i.e. they test(ed) $H_0': \theta \leqq \theta_1$, against $H_1': \theta > \theta_1$ (θ_0 acting as an idealization of H_0'). Now the critical region $w_{*'}$ which is minimax in the situation described by adding a prime to each H in the definition of L has two properties: (i) $P_{\theta_0}(w_{*'}) < \alpha = l_1/(l_0 + l_1)$, depending on θ_1, (ii) $P_{\theta_0}(w_{*'})$ is smaller with more powerful test families $\{w_t\}$. Both effects (subsisting with loss functions allowing for indifference zones) indicate that fixing $P_\theta(w_*)$ upon a constant level for such "idealized null hypotheses" as $\theta = \theta_0$ may be a questionable procedure. (Received January 28, 1957.)

26. **A Note on Fluctuations of Telephone Traffic,** VÁCLAV EDVARD BENEŠ, Bell Telephone Laboratories, (By Title).

Let $N(t)$ be defined as the number of calls in progress in a simple telephone exchange model characterized by unlimited call capacity, a general probability density of holding-time, and randomly arriving calls. A formula, due to Riordan, for the generating function of the transition probabilities of $N(t)$ is proved. From the generating function, expressions for the covariance function of $N(t)$ and for the spectral density of $N(t)$ are determined.

It is noted that the distributions of $N(t)$ are completely specified by the covariance function, if $N(t)$ is defined as above. (Received February 4, 1957.)

27. Randomization Procedures for the Estimation of Cross-Spectral Density Functions, A. E. Garratt, Virginia Polytechnic Institute, (By Title).

The cross-spectral density function may be estimated by

$$\Phi_{xy}^{*}(\omega_k) = \sum_{u=1}^{n}\{x(t_u)y(t_u + k_u\Delta t)G_1(k_u) + iX(t_u)y(t_u + m_u\Delta t)G_2(m_u)\}$$

where the k_u are independently distributed according to $p_1(k)$, $k = -r, \ldots, r$; where the m_u are similarly distributed according to $p_2(m)$ and are independent of the k_u; and where $G_1(k_u)$ and $G_2(m_u)$ are arbitrary weight functions.

It is shown that the expectation of the estimator depends on the products $p_1(k)G_1(k)$ and $p_2(m)G_2(m)$, whereas the variance of the estimator depends specifically on $p_1(k)$ and $p_2(m)$. Various specifications of the products $p_1(k)G_1(k)$ and $p_2(m)G_2(m)$ and of the probability distributions $p_1(k)$ and $p_2(m)$ are considered which provide estimators with certain optimum properties. (Received February 8, 1957.)

28. Fréchet Differentiable Functional Estimates, Gopinath Kallianpur, Michigan State University, (introduced by Morris Skibinsky).

Suppose $f_\theta(x)$ is a probability density over the finite range (a, b) which is independent of the unknown θ to be estimated. Let $\phi_n(x)$ denote an empirical density function (defined in the paper) of a sample of size n from the given population. Let G be a class of functionals over the Banach space L_1 satisfying the following conditions: (i) G possesses Fréchet differentials of the first two orders at the "true point" f_θ. If $g_1[f_\theta\,;x]$ and $g_2[f_\theta\,;x\,,y]$ are the Fréchet derivatives of the first and second order at f_θ, (ii) $g_1[f_\theta\,;x]$ is a continuous function of x which is not zero over a set of positive measure, (iii) $|\,g_2[f_\theta\,;x\,,y]\,| \leqq A < \infty$, A being independent of x and y. (iv) $G[f_\theta(x)] = \theta$ ("Fisher consistency"). Then assuming regularity conditions which validate differentiation with respect to θ, etc., and assuming $E_\theta(g_1[f_\theta\,;x]) = 0$ without loss of generality, it is shown that $\sqrt{n}\,[G[\phi_n] - \theta\}$ is asymptotically normally distributed with zero mean and asymptotic variance $E_\theta(g_1^2[f_\theta\,;x])$ which satisfies Fisher's inequality $E_\theta(g_1^2[f_\theta\,;x]) \geqq \{E_\theta[(\partial \log f_\theta)/\partial\theta]^2\}^{-1}$. An earlier paper by C. R. Rao and the author (Sankhyā, 1955) discusses similar problems for functionals of the empirical c.d.f. (Work done under ONR project at Columbia University.) (Received February 14, 1957.)

29. The Efficiency of Nonparametric Tests, Gottfried E. Noether, Boston University.

Given two tests of the same hypothesis and the same significance level. If for the same power with respect to the same alternative one requires a sample of size n_1 and the other a sample of size n_2, the relative efficiency of the second test with respect to the first test is given by the ratio n_1/n_2. The paper surveys existing results on the relative efficiency of important nonparametric tests with respect to corresponding parametric as well as other nonparametric test procedures. In particular, the following problems are considered: one-sample and paired comparison tests, two-sample tests, analysis of variance tests, tests of independence and regression, goodness of fit tests. As a general conclusion, it can be said that the employment of the more efficient nonparametric methods instead of the customary parametric methods rarely involves an appreciable loss of information, but may lead to a considerable gain. (Received March 1, 1957.)

30. **On a Problem in Abelian Groups and the Construction of Fractionally Replicated Designs,** R. C. BOSE, University of North Carolina AND R. C. BURTON, National Bureau of Standards.

Consider an Abelian group of order s^n, generated by n letters $A_1, A_2, \ldots, A_n$, with the relations $A_1^s = A_2^s = \ldots = A_n^s = I$, where I is the identity and s is a prime. If $G = A_1^{x_1}A_2^{x_2} \ldots A_n^{x_n}$ is any element of the group, then the number of non-zero exponents x_i may be called the length of G. Given an integer $r < n$, the problem is to find a subgroup of order s^r, generated by r independent elements $G_i = A_1^{x_{i1}}A_2^{x_{i2}} \ldots A_n^{x_{in}}$ such that the minimum length of the elements in the subgroup (except the length of the unit element) is greater than or equal to k. Consider the finite projective space $PG(r-1, s)$. To any point $x = (x_1, x_2, \ldots, x_r)$ of this space, assign a non-negative integer m, which may be considered the measure of x, in such a way that the total measure for the space is n. To a point of measure m associate m different letters chosen out of $A_1, A_2, \ldots, A_n$, each of these letters being assigned to one and only one point. Let $G_i = A_1^{x_{i1}}A_2^{x_{i2}} \ldots A_n^{x_{in}}$ where x_{ij} is the ith coordinate of the point to which A_j is associated. It is proved that the length of the element $G_1^{\lambda_1}G_2^{\lambda_2} \ldots G_r^{\lambda_r}$ is the measure of the set of points not lying on the linear space $\lambda_1x_1 + \lambda_2x_2 + \ldots + \lambda_rx_r = 0$. For example let $n = 10$, $r = 4$, $s = 3$. We can find exactly 10 points on an unruled quadric in $PG(3, 3)$. If we take the corresponding subgroup as the fundamental identity for generating a $\frac{1}{3^4}$ fraction in a factorial design with 10 factors, then all the aliases of a main effect will have five or more factors, and all the aliases of two factor interaction will have four or more factors. (Received January 21, 1957.)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Professor Felix Bernstein, the founder and director Emeritus of the Institute of Mathematical Statistics, University of Goettingen, Germany, died December 3, 1956 in Zuerich, Switzerland. Professor Bernstein was also a member of the International Statistics Institute, a fellow of the Royal Statistics Society, a fellow of the AAAS, and was professor of biometrics, New York University from 1936–1945. In 1950 he was American Fulbright professor at the Institute of Statistics, Rome, Italy.

Dr. Robert M. Blumenthal has been appointed to an instructorship at the University of Washington.

Glenn L. Burrows has been appointed Staff Statistician at the Knolls Atomic Power Laboratory, Schenectady, New York.

Victor Chew resigned on February 1, 1957 from the position of Assistant Professor of Statistics, University of Florida, to become Asst. Statistician at Institute of Statistics at Raleigh, North Carolina, and do work towards a Ph.D. in experimental statistics.

Professor Kai Lai Chung, on leave from Syracuse University, is a Visiting Professor at the University of Chicago during 1956–57.

George E. Ferris is now with the Statistics Department of General Foods' Corporation in Hoboken, New Jersey.

Alfred E. Garratt is now Assistant Professor of Statistics at Virginia Polytechnic Institute.

Ferdinand Lemus, formerly with the Experimental Design and Statistical Analysis Group of Westinghouse Electric Corp., East Pittsburgh, Pa., is now employed as a Statistical Engineer with the Avon Lake Experimental Station, B. F. Goodrich Chemical Company, Avon Lake, Ohio.

Gisiro Maruyama has been appointed Professor at Kyusyu University, Faculty of Science, Fukuoka, Japan.

W. Jay Merrill, Jr., received his Ph.D. from Ohio State University in December, 1956.

Robert Mirsky, who for the past six years has been with the Cornell Aeronautical Laboratory, has joined the General Electric Company as an operations research analyst in their Engineering Operation at 3198 Chestnut Street, Philadelphia, Pa.

Dr. Mervin E. Muller is on leave of absence from the Scientific Computing Center, International Business Machines Corp., New York, in order to accept a position as Research Associate in Mathematics, Department of Mathematics, Princeton University.

A. Carl Nelson, Jr., formerly an instructor in mathematics at the University of Delaware, Newark, Delaware, is now a statistician for Westinghouse Atomic Power Division, Pittsburgh, Pa.

Joseph S. Rhodes has been appointed manager of operations research of Atlas Powder Company, a new post in its Economic Evaluation Department.

Ronald W. Shephard has been appointed a professor in the Engineering Department, University of California, Berkeley, California.

George W. Snedecor has gone to North Carolina State College as a visiting professor in the Institute of Statistics, stationed with the Department of Experimental Statistics from January until September 1957. He is engaged in consulting and teaching statistical methods.

Dan Teichroew has been transferred by the National Cash Register Company to its Electronics Division in Hawthorne, California, to form a Business Systems Analysis Section in the Product Development Department.

Dr. Alan E. Treloar is currently on leave from the University of Minnesota and serving as Director of Research for the American Hospital Association in Chicago.

Frank H. Trinkl has joined The Ramo-Wooldridge Corporation as a Member of the Technical Staff, of the Computer Systems Division.

William H. Williams has been chosen to receive The George W. Snedecor Award in Statistics for 1957, by vote of the graduate faculty in statistics at Iowa State College. The award is given annually to the person judged to be most outstanding among those students at the college working toward a Ph.D. or joint Ph.D. in statistics who are expected to graduate within a specific time interval; it consists of a year's membership in the Institute of Mathematical Statistics together with a subscription to its *Annals*.

Mrs. Pearl A. Van Natta has, since September 25, 1956, been employed by The Denver Research Institute, Division of Physics, as a Research Mathematician.

Use of High Speed Computers by IMS Members*

A questionnaire was mailed to all members of the IMS in 1956 to determine the present status of the use of high speed computers by the membership and to ascertain the kind of information desired regarding these machines. Only 263 members answered the questionnaire: 119 from colleges and universities, 90 from industry and private consulting and 54 from government. As a result of the interest in having sessions on these computers at meetings of the IMS, sessions were cosponsored at both the Seattle and Detroit meetings in 1956, one session was held at the Washington, D. C. meeting in 1957 and sessions are being planned for the Atlantic City meeting. The members seemed to be about equally divided as to topics for papers at these meetings.

Eighty-seven per cent (87%) of those replying favored an expository article on the subject for the Annals. The content of such an article should emphasize a glossary of terms useful to statisticians, description of major machines and how they differ and examples of the use of the machines for statistical purposes. Only 53% favored a regular section on computing and 23% opposed such a section, 24% being indifferent. Information regarding these machines can be obtained from other journals, but many of the members were unfamiliar with either the Association for Computing Machinery (54% of those replying) and Mathematical Tables and other Aids to Computation (40% of those replying). Of those replying, only 55% had used these computers; of those who had used the computers, 56% had done some programming.

Of those who had used a high speed computer, 20% had used it for empirical sampling; 24% for Monte Carlo; 36% for data reduction; 36% for table preparation; 54% for mathematical problems; and, 62% for data analysis. Forty-eight per cent (48%) had used it for analysis of variance or regression and correlation analysis. Other types of analysis mentioned were matrix inversion or multiplication, computation of means and standard deviations, time series and spectral analysis, multivariate analysis including discriminant functions, factor analysis, contingency tables and χ^2, linear programming and inventory control, solution of equations, numerical integration, solutions of econometric models, tests of significance, distribution theory, complicated confidence limits, differential and integral equations, fitting frequency curves, Boolean algebraic equations, actuarial formulas, non-linear equations, acceptance-rejection methods, stress analysis, maxima of functions and control charts. The committee was impressed by the infrequent use of these machines for solving problems in mathematical statistics and econometrics.

* Prepared by R. L. Anderson, Chairman of IMS Committee on High Speed Machines.

University of Michigan Program in Mathematical Statistics and Probability

The University of Michigan provides training in mathematical statistics and probability, (a) leading to the master's degree, (b) leading to the doctorate. Courses in mathematical statistics and probability are conducted by staff members (including H. C. Carver, A. B. Clarke, A. H. Copeland, C. C. Craig, D. A. Darling, P. S. Dwyer, J. G. Wendel, Oscar Wesler). Teaching fellowships and assistantships are also available. Applications for these should be sent to Professor T. H. Hildebrandt of the Department of Mathematics. Additional information may be obtained by writing Professor C. C. Craig or Professor P. S. Dwyer of the Department of Mathematics.

Summer Statistical Seminar

A two-week summer statistical seminar will be held at the Endicott House in Dedham, Massachusetts, beginning July 29, 1957. The first week will be devoted to the general topic of time series with emphasis on turbulence, aeronautics, ship motion, and communication, under the chairmanship of Leo Tick of New York University.

The program for the second week will include business applications, reliability, and data reduction topics, as part of the general discussion of the impact of computers on statistical problems. Dr. Max Woodbury is chairman of this program.

Further information can be obtained from Dr. M. E. Terry, Bell Telephone Laboratories, Murray Hill, New Jersey, or the secretary, Dr. Geoffrey Beall, Gillette Safety Razor Company, Boston, Massachusetts.

University of Chicago Department of Statistics

The department of statistics at the University of Chicago, known since its organization in 1949 as the Committee on Statistics, is now called the Department of Statistics. The name was changed from Committee to Department in order to avoid confusion about the nature and status of the organization. Leonard J. Savage, who has been Acting Chairman of the Department this year, has accepted a regular appointment as Chairman beginning March 1, 1957. He succeeds W. Allen Wallis, who now is Dean of the School of Business though he continues as a member of this Department. Other members of the Statistics faculty are K. A. Brownlee, Kai Lai Chung, Sudhish G. Ghurye, Leo A. Goodman, William Kruskal, John W. Pratt, Harry V. Roberts, and David L. Wallace.

Assistance for Travel to International Congress of Mathematicians

Funds will be made available by the National Science Foundation, aided by grants from industry, to provide travel assistance for a limited number of mathematicians attending the International Congress of Mathematicians in Edinburgh, August 14–21, 1958. Grants will be made on the basis of lists prepared by the various mathematical societies. Anyone who wishes to apply through the Institute of Mathematical Statistics should notify its Secretary (see inside front cover for address) not later than October 1, giving the following information. (1) His address; (2) a statement whether he intends to present a paper and if so, whether this is by invitation; if possible the subject or title of the paper should be given; (3) a description of other travel funds, if any, available to him. The applicant should *also* write directly to the National Science Foundation, Washington 25, D. C., requesting an application form for foreign travel grants, and should return the completed form to the National Science Foundation.

DOCTORAL DISSERTATIONS IN STATISTICS, 1956

Listed below are the doctorates conferred during the year 1956 in the United States and Canada for which the dissertations were written on topics in statistics or related fields. The university, major subject, and the title of the dissertation are given in each case. Readers are invited to notify the Editor of any omissions from this list.

John L. Bagg, Michigan State University, major in mathematical statistics, "A Probability Model for Theory of Organization of Groups with Multi-valued Relations between Persons."

Anatole Beck, Yale, major in mathematics, "On the Random Ergodic Theorem."

Robert Blumenthal, Cornell, major in probability, "An Extended Markov Property."

R. V. S. Chacon, Syracuse, major in mathematics, "Some Theorems on Continuous Parameter Markov Chains."

Richard Garth Cornell, Virginia Polytechnic Institute, major in statistics, "A New Estimation Procedure for Linear Combinations of Exponentials."

Arthur P. Dempster, Princeton, major in statistics, "The Two-Sample Multivariate Problem in the Degenerate Case."

B. J. Derwort, St. Louis, major in statistics, "An Extension of the Theory of Cumulative Frequency Functions."

Olive Jean Dunn, California (Los Angeles), major in statistics, "Estimation Problems for Dependent Regression."

Sylvain Ehrenfeld, Columbia, major in mathematical statistics, "Complete Class Theorems in Design of Experiments; Part I: Complete Class Theorems

in Experimental Design; Part II: On the Efficiency of Experimental Design."

Alvin Vincent Fend, Illinois, major in statistics, "Unbiased Estimation and Admissibility and the Treatment of Ties in the Sign Test."

Thomas Shelburne Ferguson, California (Berkeley), major in statistics, Part I: "On the Existence of Linear Regression in Linear Structural Relations;" Part II: "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities."

Aubyn Freed, Illinois, major in mathematics, "On the Ergodic Theorem in Dynamical Systems with Variant Measure."

Donald A. Gardiner, North Carolina State College, major in statistics, "Some Third Order Rotatable Designs."

Donald P. Gaver, Princeton, major in probability, "Some Results in the Theory of Queques."

David G. Gosslee, North Carolina State College, major in statistics, "The Level of Significance and Power of the Unweighted Means' Test."

Shanti Swarup Gupta, North Carolina, major in statistics, "On a Decision Rule for a Problem in Ranking Means."

Bertram W. Haines, Johns Hopkins, major in biostatistics, "Some Procedures of Selecting Records for Retirement."

Lester LaVerne Helms, Purdue, major in mathematical statistics, "Convergence Properties of Martingales Indexed by Directed Sets."

G. Ronald Herd, Iowa State College, major in statistics, "Estimation of the Parameters of a Population from a Multi-Censored Sample."

William Gerow Howe, North Carolina, major in statistics, "Some Contributions to Factor Analysis."

Milton V. Johns, Columbia, major in mathematical statistics, "Contributions to the Theory of Empirical Bayes Procedures in Statistics."

Eugene Arthur Johnson, Minnesota, major in biostatistics, "On the Problems of Errors Associated with Linear Regression."

Marvin A. Kastenbaum, North Carolina State College, major in statistics, "Analysis of Frequency Data in Multiway Contingency Tables."

Therese M. Kelleher, North Carolina State College, major in statistics, "Analysis and Interpretation of Variation in Inbred Lines and F_1 Crosses in Corn."

A. R. Khalil, North Carolina State College, major in statistics, "Joint Interpretation of Heterosis and Genetic Variance in Two Corn Varieties and Their Crosses."

Orval M. Klose, Washington, major in statistics, "Topics in Distribution-Free Statistics."

Clyde Young Kramer, Virginia Polytechnic Institute, major in statistics, "Factorial Treatments in Incomplete Block Designs."

Thomas E. Kurtz, Princeton, major in statistics, "An Extension of a Multiple Comparisons Procedure."

E. J. Lytle, Jr., Florida, major in statistics, "The Determination of Some Distributions for Which the Mid-range is an Efficient Estimator of the Mean."

John Hans MacKay, North Carolina, major in statistics, "On the Efficiency of Certain Tests for 2 x 2 Tables."

Thomas A. Magness, California (Los Angeles), major in statistics, "The Use of Cumulants in the Theory and Applications of Stochastic Processes."

Carl E. Marshall, Iowa State College, major in statistics, "Cost Control of Sample Surveys by Two-Step Designs."

Samuel T. Mayo, Minnesota, major in educational psychology, "Some Designs for the Collection of and Methods for the Analysis of Enumerative Data, with Special Applications to the Follow-up of Education Graduates."

F. S. McFeely, Virginia Polytechnic Institute, major in statistics, "Decision Procedures for the Comparison of Exponential and Geometric Populations."

Dale M. Mesner, Michigan State University, major in mathematical statistics, "An Investigation of Certain Combinatorial Properties of Partially Balanced Incomplete Block Experimental Designs and Association Schemes, with a Detailed Study of Designs of Latin Square and Related Types."

Irwin Miller, Virginia Polytechnic Institute, major in statistics, "Tests of Hypotheses Involving Desirability Relations and Some Distribution Theory Connected with Gaussian Processes."

Sujit Kumar Mitra, North Carolina, major in statistics, "Contributions to the Statistical Analysis of Categorical Data."

Gilbert I. Paul, North Carolina State College, major in statistics, "A Method of Estimating Epistatic Variance in Random Mating Populations."

W. E. Perrault, St. Louis, major in statistics, "Contribution-free Population Comparisons."

John Winsor Pratt, Stanford, major in statistics, "Some Results in the Decision Theory of One-Parameter Multivariate Polya Type Distributions."

Ronald Pyke, Washington, major in statistics, "On One-Sided Distribution-Free Statistics."

Roy Radner, Chicago, major in statistics, "Team Decision Problems."

Mushfequr Rahman, McGill, major in mathematics, "A statistical Problem in the Geometry of Numbers (Star-shaped Domains of Quadratic and Hexagonal Symmetry)."

W. L. Roach, Jr., Oregon, major in statistics, "The Application of the Exponential Distribution to a Truncated Stochastic Process."

Joan Raup Rosenblatt, North Carolina, major in statistics, "On a Class of Non-parametric Tests."

Anadi Ranjan Roy, Stanford, major in statistics, "On Chi Square Statistics with Variable Intervals."

Thomas Solon Russell, Virginia Polytechnic Institute, major in statistics, "Estimation of Individual Variations in an Unreplicated Two-Way Classification."

Jagdish Sharan Rustagi, Stanford, major in statistics, "On Minimizing and Maximizing a Certain Integral with Statistical Applications."

Jerome Sacks, Cornell, major in statistics, "Asymptotic Distribution of Stochastic Approximation Procedures."

Sam Cundiff Saunders, Washington, major in statistics, "Sequential and Randomized Distribution-Free Tolerance Limits."

Daniel E. W. Schumann, Virginia Polytechnic Institute, major in statistics, "The Comparison of the Sensitivities of Experiments Using Different Scales of Measurement."

Arthur Shapiro, California (Berkeley), major in statistics, "Some Conditions for the Existence of Similar Regions."

Robert Tynes Smith III, George Washington, major in mathematical statistics, "A Stochastic Model for Economic Time Series."

Andrew Sterrett, Pittsburgh, major in statistics, "An Efficient Method for the Detection of Defective Members of Large Populations."

Hale Caterson Sweeny, Virginia Polytechnic Institute, major in statistics, "Some Results on Experimental Designs When the Usual Assumptions are Invalid."

Maurice M. Tatsuoka, Harvard, major in educational measurements, "Joint Probability of Membership and Success in a Group: An Index which Combines the Information from Discriminant and Regression Analyses as Applied to the Guidance Problem."

James G. C. Templeton, Princeton, major in statistics, "A Test for Detecting Single Cell Disturbances in Contingency Tables."

Aram Thomasian, California (Berkeley), major in statistics, "On the Magnitude of the Sum of Error Probabilities."

Hale F. Trotter, Princeton, major in probability, "Convergence of Semi-Groups of Operators."

Donald Robert Truax, Stanford, major in statistics, "Multi-Decision Problems for the Multivariate Exponential Family."

Henry Tucker, North Carolina State College, major in statistics, "Sampling for Agricultural Price Statistics."

Irving Weiss, Stanford, major in statistics, "Limiting Distributions in Some Occupancy Problems."

Oscar Wesler, Stanford, major in statistics, "A Modified Minimax Principle."

John Wesley Wilkinson, North Carolina, major in statistics, "Analysis of Paired Comparison Designs with Incomplete Repetitions."

Myron Johnson Willis, Purdue, major in mathematical statistics, "Exponential Regression."

J. W. Woll, Jr., Princeton, major in mathematics, "Homogeneous Stochastic Processes."

New Members

The following persons have been elected to membership in the Institute

November 8, 1956 to February 5, 1957

Bacon, Dr. Ralph H., Ph.D. (New York Univ.), Physicist, General Precision Lab., Pleasantville, New York.

Brooks, W. Douglas, M.Ed. (Harvard), Research Psychologist, Educational Research Corporation, 10 Craigie St., Cambridge, Mass., *131 Broadway, Arlington, Massachusetts.*

Camacho Dias, Antonio, Diplomado en Estadistios (Escuela de Estadistics, Madrid), Tecnico de la Comision Nacional de Productividad y Colaborador del Instituto de Investigaciones Estadisticas, Velazquez 47, Madrid, Spain, *Escosure 5, Madrid, Spain.*

Crowson, Henry L., M.S. (Univ. of Florida), Graduate Assistant, Univ. of Florida, *P. O. Box 3013 Univ. Station, Gainesville, Florida.*

D'Andrea Du Bois, N. S. Jr., M.S. (Howard Univ.), Grad. Student, Univ. of California, Statistical Department, *1407 Hearst St., Berkeley 2, California.*

Draper, Norman R., B.A. (Cambridge, England), Grad. Asst., Statistics Dept., Univ. of North Carolina, Chapel Hill, N. C.

Dunlap, Paul R., M.Ed. (Penn. State Univ.), Sr. Reliability Analyst, RAD-Avco Mfg. Corp., 20 So. Union St., Lawrence, Mass., *3 Kingfisher Road, Tewksbury, Massachusetts.*

Farrell, Roger H., M.S. (Univ. of Chicago), Research Asst. Digital Computer Laboratory, University of Illinois, Urbana, Illinois, *1022 W. Daniel St., Champaign, Illinois.*

Fleisher, Harold, Ph.D. (Case Inst. of Tech.), Mgr., Research Communications, International Business Machines Corp., *P. O. Box 390, Poughkeepsie, New York.*

Fricke, Theresa A., M.S. (Purdue Univ.), Statistician, B. F. Goodrich Chemical Company, Experimental Station, Avon Lake, Ohio, *1484½ E. St., Lorain, Ohio.*

Gillis, Sister Catherine Josephine, A.M. (Boston Univ.), Asst. Prof. Dept. of Mathematics, Emmanuel College, 400 The Fenway, Boston 15, Mass.

Good, I. J., Ph.D. (Cambridge Univ.), Sen. Principal Scientific Officer, Royal Naval Scientific Service, Government Communications Hdqts., Priors Road, Cheltenham, England, *25 Scott House, Princess Elizabeth Way, Cheltenham, England.*

Haley, Lawrence B., M.S. (Alabama Polytechnic Inst.), Mathematical Statistician, Army Ballistic Missile Agency, Huntsville, Alabama, *1704 La Grande, Huntsville, Alabama.*

Holmes, D. S., M.S. (Purdue Univ.), Quality Control Engineer, General Electric Co., Bldg. 36, 1 River Road, Schenectady 5, New York.

Jaynes, Lt. William E., Ph.D. (Ohio State Univ.), Res. Psychologist (MSC), Psychology Dept. Army Medical Research Laboratory, Fort Knox, Kentucky.

Jones, N. F., LL.B. (New York Univ.), Asso. Actuary, The Prudential Ins. Co. of America, Planning and Development Dept., P.O. Drawer 594, Newark 1, N. J.

Kudo, Akio, M.Sc. (Tokyo Univ.), Asst., Mathematical Inst., Faculty of Science, Kyushu Univ., Fukuoka, Japan.

Lingappaiah, G. S., M.Sc. (Mysore Univ., India), Senior Lecturer in Stat. (India); Res. Asst. (Stanford Univ.); Dept. of Stat., University of Madras, Madras 5, India; Dept. of Stat., Stanford Univ., Stanford, Calif., *P.O. Box 617, Stanford, California.*

Marshall, Albert W., (member designated by Univ. of Washington), Lab. of Stat. Research, Dept. of Mathematics, Univ. of Washington, Seattle 5, Washington.

Moriarity, John J., (member designated by Purdue Univ.), Statistical Lab. Purdue Univ., Lafayette, Indiana.

Ockelmann, Erich, Abteilungsleiter fur Marktforschung (Handlungsvollmacht), Carl Gabler, Werbegesellschaft mbH., Munchen 2, Karlsplatz 13, Germany, *Furstenriederstrasse 100/III, Munchen 42, Germany.*

Quade, Dana E., (member designated by Univ. of N. C.), Dept. of Statistics, Univ. of N. C., Chapel Hill, N. C.

Reich, Edgar, Ph.D. (Univ. of California, Los Angeles), Asst. Professor Math., University of Minnesota, Institute of Technology, Minneapolis 14, Minn.

Seguchi, Tsunetami, M.Sc. (Kyushu Univ.), Res. Worker, Mathematical Institute, Faculty of Science, Kyushu Univ., Fukuoka, Japan.

Stevens, Martin, B.S. (Temple Univ.), Reliability Engineer, General Elec. Co. (MOSD), 3198 Chestnut Street, Philadelphia 4, Pa., *1744 Hawthorne Ave., Havertown, Pa.*

Temming, Dr. Heinz, doctor rerum naturalium (Univ. Berlin), Technical director, Peter Temming Aktiengesellschaft, Gluckstadt near Hamburg, Germany.

Thionet, Pierre, Professeur agrégé de l'Universite (Paris), Administrateur à l'Institut National de la Statistique et des Etudes Economiques, à Paris, Prete au Ministare des Finances, Service des Etudes Economiques et Financires, *36 rue de Dunkerque, Paris (X^e) France.*

Washio, Yasutoshi, M.Sc. (Kyushu Univ.), Asst., Mathematical Inst., Faculty of Science, Kyushu Univ., Fukuoka, Japan.

Weckwerth, Vernon E., B.S. (Univ. of Minnesota), Inst. and Res. Fellow, Student, Univ. of Minnesota, Minneapolis 14, Minn., 1612 Huron St., St. Paul 13, Minnesota.

Wheeler, T/Sgt. R. E., U. S. Air Force, AF17261790, Hq. Sq. Sec., Hq. AAC, Box 409, APO 942, Seattle, Washington.

Wijsman, Robert A., Ph.D. (Univ. of Calif., Berkeley), Acting Asst. Prof., Statistics Department, University of California, Berkeley, California.

Yarborough, (Miss) Leone, B.S. (Oklahoma A. and M. College), Student, Statistical Laboratory, Oklahoma A. and M. College, Stillwater, Oklahoma.

REPORT OF THE WASHINGTON, D. C. MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The 1957 Eastern Regional Meeting, seventy-second meeting of the Institute of Mathematical Statistics, was held in Washington, D. C., on March 7–9, 1957, in conjunction with the Biometric Society (Eastern North American Region).

The following 145 members of the Institute registered for the meeting:

R. L. Anderson, T. W. Anderson, James B. Bartoo, Joseph Berkson, Julius R. Blum, R. C. Bose, G. E. P. Box, R. A. Bradley, A. E. Brandt, Ben Buchbinder, Bradley Bucher, Glenn L. Burrows, John A. Carpenter, Mavis B. Carroll, Richard L. Carter, Victor Chew, K. L. Chung, Joseph L. Ciminera, Willard H. Clatworthy, A. C. Cohen, Jr., Ted Colton, Wm. I. Commins, W. S. Connor, Jerome Cornfield, E. L. Cox, Gertrude Cox, Elliot Cramer, Jonas M. Dalton, Willis Davis, Reed Dawson, Besse B. Day, Francis R. DelPriore, Earl L. Diamond, Acheson J. Duncan, David B. Duncan, Arthur M. Dutton, Churchill Eisenhart, Henry Ellner, Benjamin Epstein, William B. Fetters, Clarence Fine, Spencer M. Free, John E. Freund, Fred Frishman, Donald A. Gardiner, A. E. Garratt, John J. Gart, Seymour Geisser, Dorothy M. Gilford, R. Gnanadesikan, Mina H. Gourary, Franklin A. Graybill, B. G. Greenberg, Samuel W. Greenhouse, Joseph A. Greenwood, T. N. E. Greville, Shanti S. Gupta, John Gurland, J. S. Hagan, Max Halperin, Bernard Harris, Boyd Harshbarger, Wassily Hoeffding, W. H. Horton, Harold Hotelling, Wm. G. Howard, David C. Hurst, Frederick V. Hurst, Jr., T. A. Jeeves, G. Kallianpur, Leo Katz, Mortimer B. Keats, George H. Kennedy, A. W. Kimball, Carl F. Kossack, Charles Kraft, Clyde Y. Kramer, Morton Kupperman, Lonnie Lasman, Fred C. Leone, Alfred Lieberman, Gilbert Lieberman, Julius Lieblein, Ardie Lubin, Eugene Lukacs, J. A. McFadden, G. T. McLoughlin, Clifford J. Maloney, Arthur S. Marthens, Paul Meier, William Mendenhall, Dale M. Mesner, Herbert A. Meyer, Donald F. Morrison, Jack Moshman, Mervin Muller, Mary G. Natrella, George Nicholson, Gottfried E. Noether, Junjiro Ogawa, Carl R. Ohman, Ingram Olkin, K. M. Patwary, John F. Pauls, B. E. Phillips, Lila Knudsen Randolph, Wyman Richardson, Donald L. Richter, David Rosenblatt, Harry M. Rosenblatt, Joan R. Rosenblatt, M. Rosenblatt, S. N. Roy, Jagdish S. Rustagi, Rose Sachs, William H. Sammons, Daniel E. Sands, Marvin Schneiderman, Norman Severo, Oliver A. Shaw, Walt R. Simmons, Rosedith Sitgreaves, Morris Skibinsky, Romuald Slimak, Jean F. Smolak, Milton Sobel, Paul N. Somerville, Fannie A. Stinson, H. C. Sweeny, Zen Szatrowski, Robert J. Taylor, Henry Teicher,

M. E. Terry, Malcolm E. Turner, H. Robert van der Vaart, Irving Weiss, Kathleen White, M. B. Wilk, John W. Wilkinson, Evan J. Williams, R. Lowell Wine, Charles W. Wright, W. J. Youden, Samuel Zahl, Marvin Zelen.

The program was as follows:

THURSDAY, MARCH 7, 1957

9:45–10:00 a.m. Welcome from Catholic University
10:00–12:00 noon. Sessions A and B

Session A: Sample Survey Methodology

Chairman: BOYD HARSHBARGER, Virginia Polytechnic Institute

Papers:
1. *A Sampling Study of Sources of Information for Farm Families in Virginia*, LOWELL WINE, Virginia Polytechnic Institute
2. *A New Approach to General Purpose Sampling*, CARL KOSSACK, Purdue University
3. *Recent Experiences with Area Sampling for Agricultural Statistics*, R. E. VICKERY, Agricultural Marketing Service

Session B: Contributed Papers I (I.M.S.)

Chairman: JOAN R. ROSENBLATT, National Bureau of Standards

Papers:
1. *Synchronization of Trajectory Images of Ballistic Missiles and the Timing Record of the Ground Telemetry Recording System*, HARRY P. HARTKEMEIER, Stanford University (Introduced by Paul R. Rider)
2. *A Further Contribution to the Theory of Systematic Statistics*, JUNJIRO OGAWA, University of North Carolina
3. *Multivariate Analysis of Variance*, S. N. ROY, University of North Carolina
4. *Confidence Bounds Associated with Multivariate Analysis of Variance*, S. N. ROY AND R. GNANADESHIKAN, University of North Carolina.
5. *On Statistics Independent of a Sufficient Statistic*, EVAN J. WILLIAMS, North Carolina State College
6. *On a Problem in Abelian Groups and the Construction of Fractionally Replicated Designs*, R. C. BOSE, University of North Carolina AND R. C. BURTON, National Bureau of Standards
7. *The Variance of Zero-crossing Intervals*, J. A. MCFADDEN, U. S. Naval Ordnance Laboratory, (Introduced by Gilbert Lieberman)
8. *Maximum Likelihood Estimates in a Simple Queue*, A. BRUCE CLARKE, University of Michigan, (By Title)
9. *The Joint Distribution of a Set of Sufficient Statistics for the Parameters of a Simple Telephone Exchange Model*, VÁCLAV EDVARD BENEŠ, Bell Telephone Laboratories, (By Title)
10. *An Extension of the Cramér-Rao Inequality*, JOHN J. GART, Virginia Polytechnic Institute, (By Title)
11. *A Limit Theorem and Bounds for an Optional Stopping Probability*, MORRIS SKIBINSKY, Michigan State University, (By Title)
12. *On the Mathematical Principles Underlying the Theory of the Chi-square Test*, JUNJIRO OGAWA, University of North Carolina, (By Title)
13. *A Limit Theorem of Cramér and Its Generalization*, JUNJIRO OGAWA, University of North Carolina, (By Title)
14. *Some Uses of Quasi-Ranges II*, J. T. CHU AND F. C. LEONE, Case Institute of Technology AND C. W. TOPP, Fenn College, Cleveland, Ohio

1:30–3:00 p.m. Applications of Stochastic Processes

Chairman: RALPH A. BRADLEY, Virginia Polytechnic Institute

Papers:
1. *Some Estimation Problems in Generalized Harmonic Analysis*, J. E. FREUND AND W. O. ASH, Virginia Polytechnic Institute
2. *Applications of Stochastic Process Theory to Problems in Aeronautics*, FRANKLIN W. DIEDERICH, National Advisory Committee for Aeronautics, (Langley Field)

3:00–4:00 p.m. Tea and Social Hour

4:00–6:00 p.m. Design of Experiments

Chairman: D. B. DUNCAN, University of North Carolina

Papers:
1. *An Industrial Example of Fractional Factorials*, W. H. HORTON, Westinghouse Electric Corp.
2. *Some Problems in Evolutionary Operation*, G. E. P. BOX, Princeton University
3. *Factorial Treatments in Group Divisible Incomplete Block Designs*, CLYDE Y. KRAMER, Virginia Polytechnic Institute
4. *The Structure of Incidence Matrices of Partially Balanced Incomplete Block Designs*, D. M. MESNER, Purdue University Center (Ft. Wayne)

FRIDAY, MARCH 8, 1957

9:00–11:00 a.m. Electronic Computers

Chairman: JACK MOSHMAN, Council for Economic and Industry Research

Papers:
1. *The Use of Generalized Subroutines in Statistical Calculation*, ROMUALD SLIMAK, Sperry Rand Corp.
2. *Some Uses of the I.B.M. 650 in Applied Statistics*, H. L. LUCAS, Princeton University
3. *The Mutual Troubles of Statisticians and Digital Computers*, FORMAN S. ACTON, Princeton University

Discussant: MILTON E. TERRY, Bell Telephone Laboratories

11:00–12:00 noon. Methodology in Survey of Smoking Habits

Chairman: B. G. GREENBERG, University of North Carolina

Paper:
1. *An Investigation on Smoking Habits of Individuals*, D. G. HORVITZ, G. T. FORADORI, J. MONROE, J. FLEISCHER, A. L. FINKNER, North Carolina State College

Discussant: JEROME CORNFIELD, National Institutes of Health

2:00–4:00 p.m. Contributed Papers (I.M.S. and Biometric Society)

Chairman: MAX HALPERIN, National Institutes of Health

Papers:
1. *A Rank Order Test for Trend in Correlated Means*, ARDIE LUBIN, Walter Reed Army Institute of Research, Washington, D. C.
2. *On the Stochastic Structure of Minkowski-Leontief Systems*, DAVID ROSENBLATT, American University
3. *The Use of Incomplete Block Designs for Asymmetrical Factorial Arrangements*, MARVIN ZELEN, National Bureau of Standards
4. *Extension of Some Results Given by Mitra on Statistical Analysis of Categorical Data*, EARL DIAMOND, University of North Carolina
5. *Testing of Hypotheses on a Mixture of Variates Some of Which are Continuous*

and the Rest Categorical, S. N. Roy and M. D. Moustafa, University of North Carolina

6. *Generalized Quantal Response in Biological Assay*, John Gurland, Iowa State College
7. *The Recovery of Intervariety Information*, Bradley D. Bucher, Princeton University
8. *On the Stochastic Structure of Minkowski-Leontief Systems, II*, David Rosenblatt, American University, (By Title)
9. *On the Stochastic Structure of Minkowski-Leontief Systems, III*, David Rosenblatt, American University, (By Title)
10. *On Selecting a Subset which Contains All Populations Better Than a Standard*, Shanti S. Gupta and Milton Sobel, Bell Telephone Laboratories
11. *On the Relation Between Loss Functions and Significance Levels*, H. R. van der Vaart, North Carolina State College

4:00–6:00 p.m. Recent Developments in Statistics and Probability

Chairman: Churchill Eisenhart, National Bureau of Standards

Papers:
1. *Statistical Inference*, Jerome Cornfield, National Institutes of Health
2. *The Efficiency of Non-parametric Tests*, Gottfried Noether, Boston University
3. *A Survey of the Theory of Analytic Characteristic Functions*, Eugene Lukacs, Catholic University
4. *Tests Concerning the Means of Certain Distributions*, Norman Severo, National Bureau of Standards

SATURDAY, MARCH 9, 1957

10:00–12:00 noon. Stochastic Problems in Physics (I.M.S.)

Chairman: Kai Lai Chung, University of Chicago

Papers:
1. *Applications of Stochastic Processes to Problems in Chemical Kinetics*, Elliot Montrol, University of Maryland
2. *Statistical Theory of Cascade Processes*, J. E. Moyal, Columbia University

Dorothy Morrow Gilford
Associate Secretary

PUBLICATIONS RECEIVED

Diab, M. A., *The United States Capital Position and the Structure of Its Foreign Trade* (*Contributions to Economic Analysis*), North-Holland Publishing Company, Amsterdam, P.O. Box 103, Holland, 1956, viii + 67 pp., $2.75.

Clatworthy, W. H., *Contributions on Partially Balanced Incomplete Block Designs with Two Associate Classes*, National Bureau of Standards Applied Mathematics Series 47, 70 pages, 7 tables, 45 cents. (Order from the Superintendent of Documents, Government Printing Office, Washington 25, D. C.)

Narasimham, N. V. A., *A Short Term Planning Model for India* (*Contributions to Economic Analysis*), North-Holland Publishing Company, Amsterdam, P.O. Box 103, Holland, 1956, xiii + 93 pp., $2.75.

Tinbergen, J., *Economic Policy: Principles and Design* (*Contributions to Economic Analysis*), North-Holland Publishing Company, Amsterdam, P. O. Box 103, Holland, 1956, xxviii + 276 pp., $7.00.

TRABAJOS DE ESTADISTICA

Review published by "Instituto de Investigaciones Estadísticas" of the "Consejo Superior de Investagaciones Científicas." Madrid, Spain.

Vol. VIII **CONTENTS** **Cuaderno I**

For everything in connection with works, exchanges and subscription write to Professor Sixto Ríos, Instituto de Investigaciones Estadísticas of the Consejo Superior de Investigaciones Científicas (Serrano, 123). Madrid, Spain. The Review is composed of three fascicles published three times a year (about 350 pages), and its annual price is 100 pesetas for Spain and South America and $4.00 U.S.A. for all other countries.

ESTADISTICA

Journal of the Inter American Statistical Institute

Volume XIV, No. 53 **Contents** **December 1956**

Published quarterly *Annual subscription price* $3.00 (U. S.)

INTER AMERICAN STATISTICAL INSTITUTE

Pan American Union
Washington 6, D.C.

ECONOMETRICA

Journal of the Econometric Society

Contents of Vol. 25, No. 2 - April, 1957

JOURNAL OF THE
ROYAL STATISTICAL SOCIETY

Series B (Methodological)

Vol. XVIII, No. 2, 1956

The Royal Statistical Society, 21, Bentinck Street, London, W. 1